Clinically Aware Data Summarization at the Edge for Internet of Medical Things

(Ph.D. Forum)

Rahul Krishnan Pathinarupothi
Amrita Center for Wireless Networks & Applications (AmritaWNA), Amritapuri Campus
Amrita School of Engineering, Amrita Vishwa Vidyapeetham, India
rahulkrishnan@am.amrita.edu

***Abstract*—Internet of Medical Things (IoMT) enable early detection and alerting of critical disease conditions through continuous monitoring of electro-physiological body parameters. However, large scale use, especially in rural and sparsely connected regions, poses challenges in terms of unavailability of sufficient data bandwidths and connectivity. Additionally, the physicians, who are already managing huge patient load, would be overwhelmed to take clinical decisions after viewing voluminous data pouring in from multiple IoMT devices.**

In our research, we developed a technique called Rapid Active Summarization for effective PROgnosis (RASPRO) that converts unwieldy multi-sensor time series data into summarized patient/disease specific trends in steps of progressive precision as demanded by the physician for patients personalized condition at hand and help in identifying and subsequently predictively alerting the onset of critical conditions. RASPRO generates clinically useful, yet extremely succinct, summary of a patient's medical condition represented as a series of symbols called motifs, which could be sent to remote physicians even over SMS or emerging narrow bandwidth Internet of Things (IoT) networks. We demonstrate that the diagnostic predictive power of RASPRO motifs is comparable to raw sensor data.

***Index Terms*—Internet of Medical Things, Severity detection, clinical data summarization**

I. Introduction

Wearable IoMT devices is making remote monitoring of patients widely accessible. However, large scale deployment of these devices is still a far away reality. In most of the rural and remote regions there is only intermittent connectivity to data networks along with scarcity of bandwidth. Communicating the vital parameters from IoMT devices to remotely located physicians demands development of bandwidth miserly data reduction techniques. There is a complementary challenge at the physician's end too. The physicians, who are already overloaded, would feel even more overwhelmed by the voluminous data being flooded from remote patients' sensors.

Researchers have actively looked at how to reduce the data bandwidth consumption. *Hooshmand et al.* [1] evaluate the suitability of compression techniques based on Fourier transform, DCT, and wavelets, and propose online dictionaries as a preferred technique for power-efficient transmission of biosignals. *Lee et al.* [2] demonstrate an efficient compression technique for ECG signals that could be used for real-time applications. Some of the earlier works in symbolic representation of data such as Symbolic Aggregate approXimation (SAX) *Lin et al.* [3] convert data into simpler, dimensionally reduced symbols. However, these existing data summarization, reduction, and data fusion [4] techniques do not lend themselves to segregation based on severities set by medical experts, potentially resulting in loss of clinical interpretability, thereby missing clinically relevant insights as well as resulting in clinical inconsistencies [5]. These techniques also involve computationally complex encoding-decoding algorithms that might put additional constraints on IoMT edge devices.

In our research, we have developed and evaluated a technique that we call as Rapid Active Summarization for effective PROgnosis (RASPRO). Our RASPRO technique converts unwieldy multi-sensor time series data into summarized patient/disease specific trends in steps of progressive precision as demanded by the physician for patients' personalized condition at hand and help in identifying and subsequently predictively alerting the onset of critical conditions.

II. Rapid Active Summarization for effective PROgnosis

Let us consider the raw sensor data coming in from multiple IoMT sensor devices. There are four major processing steps (see Figure 1).

Step 1: From the sensor data, we derive N different vital parameter series, VPS_1, VPS_2, ..., VPS_N. For example, in our system there are five ECG (QRS width, QTc, R-R interval, S-T elevation/depression, P-R interval) and three PPG (BP, SpO2, pulse rate) parameter series, totaling up to $N = 8$.

Step 2: Each sample value in each of the vital parameter series is converted to disease-parameter specific discrete quantized severity level symbol such as, 'A', '$A+$', '$A++$', '$A-$', '$A--$', etc., where the clinically defined normal range for a given parameter is assigned the symbol 'A', while values which are above and below normal ranges are assigned with increasing number of "$+$" and "$-$" suffixes according to the severity. Unlike typical systems (such as the one described by *Wu et al.* [6]) with fixed severity thresholds for a sensor, in RASPRO the number of severity level symbols L_{SVR} and their mapping to corresponding parameter value ranges are customized according to the patient's disease.

The output of *Step 2* are multiple series of quantized severity level symbols corresponding to different parameters.

978-1-5386-9151-9/19/$31.00 ©2019 IEEE

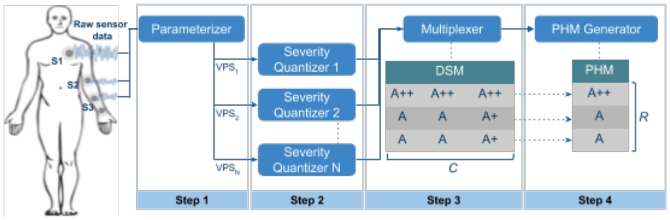

Fig. 1. The RASPRO Engine for converting raw sensor data from IoMT devices to severity summaries in the form of Disease Severity Matrices (DSMs) and Personalized Health Motifs (PHMs).

Step 3: In this step, quantized severity symbol series of all parameters relevant to the disease being monitored are grouped into a two dimensional Disease Severity Matrix (DSM) by a multiplexer. Each row R in the DSM represents one parameter, and the number of columns C represents the total number of samples in the monitoring interval.

There may be a succession of DSMs for successive monitoring intervals with a quiescent time gap of δ between them.

Step 4: In the final step of RASPRO, all the quantized severity symbols in a parametric row of DSM are temporally summarized to one *consensus symbol* using a disease-parameter specific formulation. In general, the consensus symbol captures the dominant trend in the patient data.

Figure 1 depicts an example where there are three rows in the DSM corresponding to three parameters, namely blood pressure (BP), blood oxygen saturation (SpO2) and pulse rate. Each of these parametric rows is summarized to consensus symbols (which are defined here as the most frequently occurring symbol) "$A++$", "A", and "A", respectively.

Personalized Health Motifs: Consensus symbols corresponding to all the rows in a DSM are put together in a column vector to arrive at a Personalized Health Motif (PHM) as depicted in Figure 1. Different kinds of summarization based on the sensor type, diagnostic interest, and patients' health condition are possible. Depending up on these factors, the physician might want to know the mean of values, frequency of peaks, value of highest peak, most frequent abnormality, sparsity of values etc. Accordingly, the RASPRO provides a framework to define summarization differently for various clinical requirements. A more exhaustive approach to defining PHMs to capture disease specific trends is discussed in one of our recent papers [7].

Motifs as input to Machine Learning Models: Machine learning classifiers that are domain agnostic can result in mis-classification unless trained on large and reliable enough datasets. Such datasets need to adequately represent the complete spectrum of variations observed for disease conditions. Currently, the number of diseases for which such datasets are publicly available is limited. Also, classifiers could be comparatively computationally complex to execute on an IoMT edge-device, and hence are best suited to run on the cloud to aid in automated diagnosis. However, since RASPRO motifs represent the major trends seen in the patient, it could be used as an alternative dataset to build machine learning models that could potentially run on reduced datasets.

III. Evaluation and Results

The RASPRO technique was clinically validated in two steps: (a) Physionet MIMIC II dataset [8] based testing on 83 patients for detecting and advance warning of cardiac conditions and (b) comparative analysis with another symbolic data reduced representation technique [3].

The results, which are elaborated in our recently published works [5], [9], demonstrate that RASPRO PHMs have very good predictive power, as measured by standard metrics of precision, recall, and F1-score (with values equal to 0.87, 0.83, and 0.85 respectively). Also, we show that RASPRO outperforms domain agnostic techniques such as SAX; 20-90% improvement in F1 score over bandwidths ranging from 0.2-0.75 bits/unit-time. These results are attributable to the premise that PHMs effectively capture the major abnormal trends for a first level clinical diagnosis, and suppress details that can be deferred for a subsequent detailed investigation.

IV. Conclusion and Future Work

RASPRO is a medically-aware clinical severity detection and data summarization technique for IoMT devices. This technique can be applied in the edge devices to make pervasive computing more accessible in remote and sparsely connected regions. Future direction of research includes opportunities to model and validate RASPRO in other specialties, such as neurology and endocrinology.

Acknowledgment

The author's Ph.D. thesis advisor is Dr. P Venkat Rangan. The author also thanks Sri Mata Amritanandamayi Devi, Chancellor of Amrita University, for her continuous encouragement and support towards pursuing research that has direct societal impact.

References

[1] M. Hooshmand, D. Zordan *et al.*, "Boosting the battery life of wearables for health monitoring through the compression of biosignals," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1647–1662, Oct 2017.

[2] S. Lee, J. Kim, and M. Lee, "A Real-Time ECG Data Compression and Transmission Algorithm for an e-Health Device," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 9, pp. 2448–2455, sep 2011.

[3] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*. ACM, 2003, pp. 2–11.

[4] S. Abhishek, S. Veni, and K. Narayanankutty, "Biorthogonal wavelet filters for compressed sensing ecg reconstruction," *Biomedical Signal Processing and Control*, vol. 47, pp. 183–195, 2019, cited By 0.

[5] P. Durga, R. K. Pathinarupothi *et al.*, "When less is better: A summarization technique that enhances clinical effectiveness of data," in *Proceedings of the International Conference on Digital Health*, ser. DH '18. New York, NY, USA: ACM, 2018, pp. 116–120.

[6] S.-J. Wu, R.-D. Chiang, S.-H. Chang, and W.-T. Chang, "An Interactive Telecare System Enhanced with IoT Technology," *IEEE Pervasive Computing*, vol. 16, no. 3, pp. 62–69, 2017.

[7] R. Pathinarupothi, P. Durga, and E. Rangan, "Data to diagnosis in global health: A 3P approach," *BMC Medical Informatics and Decision Making*, vol. 18, no. 1, 2018.

[8] A. L. Goldberger, L. A. Amaral *et al.*, "Physiobank, physiotoolkit, and physionet," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[9] R. Pathinarupothi, P. Durga, and E. Rangan, "IoT Based Smart Edge for Global Health: Remote Monitoring with Severity Detection and Alerts Transmission," *IEEE Internet of Things Journal*, 2018, (to appear).