Confidence Metric in Critical Systems

Minho Shin

Myongji University

1 Introduction

The advent of portable computing devices and miniature sensing devices presents many new opportunities for personal healthcare. Formerly, most medical sensing devices were used in a hospital setting under the care of trained medical and technical personnel; soon, many devices will be worn throughout a patient's daily life or installed at home and in assisted-living settings.

These devices will collect health-related data for many purposes, by patients with chronic medical conditions (such as blood-sugar sensors for diabetics), people seeking to change behavior (e.g., losing weight or quitting smoking), or athletes wishing to monitor their condition and performance. The resulting data may be used directly by the person, or shared with others: with a physician for treatment, with an insurance company for coverage, with the adult children of elderly parents, or by a coach.

Such systems have huge potential benefit to the quality of healthcare and quality of life for many people, but there are many opportunities for the sensor data to be tampered or otherwise inaccurate. Outside the hospital setting, in particular, the sensors may be applied by the patient or family members; the data may be gathered through a personal mobile device (such as a mobile phone), over a personal network (such as a wireless network at home), or over the public Internet. Therefore, the accuracy and availability of sensor data is difficult to ensure.

To evaluate the trustworthiness of medical data gathered in this manner, we need a holistic system view that inspects contributions to risk and error in medical data as it flows from the patient to the caregiver. To be useful, systems must assure that high-quality information reaches the data user – or, at least, the system must be able to express some degree of confidence in the data being presented.

Knowing the degree of confidence in data can be beneficial. First, caregivers can make more accurate decisions based on an understanding of confidence in the data. A quantifiable metric for confidence also allows for quality control of the system. A well-designed system can detect an anomaly in data, whether accidental or malicious, and alarm the caregivers for further verification.

1.1 Problem Statement

The project aims to answer the following question: *how can we assess confidence in sensor data in the context of pervasive health monitoring, and how can we*

present the confidence to user? To answer the question, we propose the following research hypothesis.

1.2 Research Hypothesis and Objectives

In this section, we propose our research hypothesis statement, list research questions, and identify the research objective for each research question.

Hypothesis. "Confidence" may be coarsely quantified and derived from a combination of several factors.

Our hypothesis is that we can express confidence in sensor data with a quantifiable metric. To test our hypothesis, we raise the following research questions.

Q1. What factors can contribute to the confidence in sensor data?
Q2. How can we quantitatively measure the overall confidence in sensor data?
Q3. How can we effectively present the confidence metric to users?

2 Confidence Metric

In pervasive health monitoring, each patient carries sensors and a *collection device*, such as personal mobile phone. The collection device gathers sensor readings from sensors and reports to the central server through available network connections. We call a set of successive reports coming from the same collection device a *sensing stream*, but as a *snapshot* of the sequence of data rather than a whole set of accumulated reports. The *confidence* in a sensing stream is the belief that the reports, especially recent and incoming ones, contains correct values for the patient's physiology. The *confidence metric* is a quantitative assessment of the degree of confidence in a sensing stream.

The confidence metric should meet the following requirements:

- The metric provides the best assessment of the system's confidence in data, given knowledge of the system configuration and architecture, training data, and the history of sensor readings.
- The metric is concise, desirably a single numeric value, but takes into account all the underlying factors that affect the data quality.
- The metric is time-dependent and represents the temporal change in data quality over time.

To measure the confidence in data, the system needs to know the ground truth. In a pervasive health-monitoring, however, system often has no access to the ground truth. Therefore, the system estimates the ground truth with supporting evidences at hand. Such evidences include the known facts about the system, equipments, and the patient; a large database of physiological data for a specific patient or a set of patients; and the (recent) history of the sensing stream.

Pervasive health monitoring is dynamic. For example, the sensors can be applied differently everyday by the patient. The data can travel through a private corporation network today but through a public Wi-Fi network tomorrow. Even the accuracy of the sensor may decay over time. It is essential for the confidence metric to capture temporal dynamics of the situation, so the changes can be reflected in the metric.

Many factors contribute to confidence in data. Since a mishap in any factor can degrade the data quality, the confidence metric should depend on each factor. Furthermore, the metric should reflect the relationship among factors. For example, if two factors are independent to each other, a change in one factor should not change the contribution of the other independent factor.

Due to vast differences between sensing streams, in terms of sensor set, architecture, and the unique physiology of a patient, we do not intend to provide an absolute metric that can precisely compare confidence metrics of different sensing streams. Instead, the metric itself provides only a coarse assessment of the confidence in data, while the relative changes over time can be a good judgment about the dynamics of the data quality.

2.1 Definition

The *confidence metric* of a sensing stream is an evidence-based quantification of the truthfulness of the recent and incoming sensor reading in the sensing stream. A sensing stream is *truthful* if the physiological data represents the *ground truth*, i.e., the physiological data we expect when we monitor the patient in the laboratory with ideal sensors. Since we cannot know the ground truth, we estimate the probabilistic distribution of the ground truth and then evaluate the truthfulness of the sensing stream against the probabilistic ground truth.

We can estimate the probabilistic ground truth based on *evidences* at hand. Such evidences include known facts (e.g., the system configuration, system architecture, or the patient's medical record), patient's physiological model (e.g., obtained from training phase), or annotated data (e.g., cryptographic footprints or certificates).

Even with the ground truth, evaluating truthfulness is not trivial because of the uncertainty involved in sensor data. First, a physiological value itself follows a probabilistic distribution. Second, the accuracy of a sensor is probabilistic. Third, threats on data quality, whether intentional or accidental, can be unpredictable. Therefore, we model confidence metric as a probability that represents our *best guess* about the truthfulness of the data, given evidences at hand.

Definition 1 (Confidence Level).
We define confidence metric at time t by the probability that the sensing stream is truthful conditioned by recent observation, or

$$\mathrm{CL}_t = P[T_t = 1 | O_t = \mathbf{d}_t] \tag{1}$$

where T_t is a indicator random variable for "the sensing stream is truthful at time t", O_t is an observation variable, and $\mathbf{d}_t$ denotes a set of recent sensor readings.

2.2 Factorization

Our approach is *divide-and-conquer*. We decompose the problem into eleven mutually independent factors, analyze each factor to derive a sub-confidence metric, and combine them to quantify the overall confidence level. In this section, we discuss the decomposition of confidence into factors.

Note that the sensing stream is truthful only when it is truthful with respect to all the factors. Formerly, the event $\{T_t = 1\}$ is equivalent to the event $\{T_t^{S_1} = 1 \wedge T_t^{S_2} = 1 \wedge \ \ldots \ \wedge T_t^{A_3} = 1\}$. Because of the independence between factors, we get

$$\mathrm{CL}_t = \mathrm{CL}_t^{S_1} \times \cdots \times \mathrm{CL}_t^{A_3} \tag{2}$$

Therefore, assessing overall confidence metric boils down to assessing the sub-confidence metric of each factor. In the following, we discuss how to derive sub-confidence metric.

2.3 Factor analysis model

In the following, we discuss three possible models for deriving sub-confidence metrics.

Black-and-white confidence model. Some factors may have a bipolar contribution to data quality; either completely keeping the data close to the ground truth or completely falsifying the data. To apply this model, the system should provide a black-and-white verification method, say $\mathsf{Verify}(m)$, which outputs 0 (when verification fails) or 1 (when verification succeeds) for the sensor data m.

Knowledge-based confidence model. For some factors, deriving confidence can be merely referencing a knowledge database that provides recommended confidence level of given factor. For example, the confidence of a specific sensor model can be looked up from a knowledge database. Building such database will require close analysis, lab experiments, or a reputation system.

Functional confidence model. A factor can affect data quality in a fine-grained mathematical way so that the data quality can be modeled by a function of measurable inputs such as the observation or known facts. Let us denote the data at time t by $\mathbf{d}_t$ and a set of known facts by F. A *confidence function* $\mathsf{CF}^X(\mathbf{d}_t, F)$ is a continuous function that outputs a sub-confidence level of factor X.

Bayesian model. When uncertainty plays a major role to the confidence, we can use probabilistic approach. In particular, we can use the well-known Bayesian Theorem to compute the sub-confidence level of factor X to get

$$P_t[T_t^X = 1 | O_t = \mathbf{d}_t] = c \cdot P_t[O_t = \mathbf{d}_t | T_t^X = 1] \cdot P_t[T_t^X = 1] \tag{3}$$

where the first probability of the right hand side is called *likelihood* and the second one called *prior probability*. The c is a constant to normalize the probability of the left hand side. The likelihood ($P_t[O_t = \mathbf{d}_t | T_t^X = 1]$) represents the probability of observing a specific data when the sensing stream is truthful. To compute the likelihood value, we need to know the probability distribution of the ground truth. We can estimate the ground truth based on analysis or heuristics or we can learn it from training data.

2.4 Presentation of Confidence Metric

As a preliminary discussion for the presentation issue, we set forth the diagnostic use of confidence metric as a starting point. The goal of diagnostic presentation of confidence metric is to help users explore the underlying reasons for a poor confidence. To that end, the system may provide sub-confidence metrics or further analytic information when needed.

It is essential to conduct a user study for discovering other presentation issues that the caregivers are concerned about.

- (Cross-time presentation) To inform the user of temporal change of confidence level and possibly its temporal trend, we consider graphical representation.
- (Smooth presentation) To minimize the confusion due to high variance in confidence level, we apply smoothing techniques such as moving average on sensor data or on the confidence factors.
- (Diagnostic presentation) To help users with further action against poor confidence levels, the system may provide individual confidence factor or further analytic information when needed.

3 Gap Analysis

Several recent studies have investigated the potential of using physiological signals such as the Electrocardiogram (ECG), Photoplethysmograph (PPG) etc. for biometric identification and verification [2,1,3,5,7,8]. The findings from the studies indicate that the reliability of the mechanism degrades with varied sensing conditions such as activity or stress levels of the sensor-wearer that cause intra-subject variance. Hence these approaches are unsuitable for the desired ongoing verification of patient identity. A related study by Jea et al., [6] employs multiple sensor modalities (heart rate, blood pressure and weight) for biometric identification of users. However we believe that the reliability of such an approach can be further improved using accelerometry and galvanic skin response as additional modalities to rationalize trends in ECG sensor data during varied conditions. A closely related study is the HUMABIO project [4] for unobtrusive multimodal biometric authentication. The project aims to use dynamic physiological user profiles to unobtrusively validate user identity within a secure area after initial authentication. The physiological sensor modalities investigated in pilot studies

of the project include ECG and EEG. Findings indicate that features extracted from the heart beat shape have high discriminative power. However the HUMABIO approach is unreliable during intense activity and other varied conditions due to the assumption of controlled conditions within the secure area. We will investigate multiple modalities of sensor data taken together for a more robust approach.

4 Conclusion

In this paper, we present our approach to measuring confidence in sensing data collected through remote monitoring. In future work, we plan to conduct factor analysis and then derive confidence metric that takes account those factors and influences of those factors to the confidence level.

5 Bibliography

References

1. F. Agrafioti and D. Hatzinakos. Fusion of ecg sources for human identification. *In proceedings of the 3rd International Symposium on Communications, Control and Signal Processing, ISCCSP 2008*, pages 1542–1547, March 2008.
2. L. Biel, O. Pettersson, L. Philipson, and P. Wide. Ecg analysis: a new approach in human identification. *Proceedings of the 16th IEEE Instrumentation and Measurement Technology Conference*, 1:557–561, 1999.
3. Y.Y. Gu, Y. Zhang, and Y.T. Zhang. A novel biometric approach in human verification by photoplethysmographic signals. *In Proceedings of the 4th Annual IEEE Conference on Information Technology Applications in Biomedicine, UK*, pages 13–14, April 2003.
4. Evangelos Bekiaris Ioannis G. Damousis, Dimitrios Tzovaras. Unobtrusive multimodal biometric authentication: The HUMABIO project concept. *EURASIP Journal on Advances in Signal Processing*, 2008.
5. Steven A. Israel, John M. Irvine, Andrew Cheng, Mark D. Wiederhold, and Brenda K. Wiederhold. Ecg to identify individuals. *Pattern Recognition*, 38:133–142, January 2005.
6. David Jea, Jason Liu, Thomas Schmid, and Mani B Srivastava. Hassle free fitness monitoring. In *Proceedings of the 2nd International Workshop on Systems and Networking Support for Healthcare and Assisted Living Environments (HealthNet)*, Jun. 2008.
7. K.N. Plataniotis, D. Hatzinakos, and J.K.M. Lee. Ecg biometric recognition without fiducial detection. *Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference*, 19:1–6, Aug 2006.
8. T.W. Shen, W.J. Tompkins, and Y.H. Hu. One-lead ecg for identity verification. *Proceedings of the 24th Annual Conference Engineering in Medicine and Biology and the Annual Fall Meeting of the Biomedical Engineering Society, EMBS/BMES Conference*, 1:62–63, 2002.