PhD Thesis Report

"Sensor Information Scoring for Decision-Aid Systems in Railway Domain", by MARCIN LENART

Report done by Dr. hab. GRÉGORY SMITS the 25^{th} of November, 2019

IRISA - Université Rennes 1
6, rue de Kerampont
CS 80518, 22305 Lannion cedex
gregory.smits@irisa.fr

In this PhD report, MARCIN LENART describes his research works financed by Thales Polska, supervised Dr. hab. ANDRZEJ BIELECKI (AGH university) and Dr. hab. MARIE-JEANNE LESOT (Sorbonne université) and advised by Dr. TEODORA PETRISOR (Thales) and Dr. ADRIAN REVAULT D'ALLONNES (Sorbonne université).

Introduction to the Addressed Problem This work addresses a crucial and topical issue that concerns the question of how to assess information quality provided by sensors. This task is particularly useful when information serve as a basis to decision making. Due to the increase of the number of sensors used in very different applicative contexts to measure physical phenomena, tools have to be defined to help domain experts understand the output of these devices and make the most of the information provided.

The goal of MARCIN LENART's work is to provide a generic approach to quality assessment of the information provided by sensors.

Document Structure and Comments The doctoral thesis is divided into 5 chapters boxed by an introduction and a conclusion also providing perspectives for future works.

The first chapter is dedicated to a reminder about basic notions on which this work relies. An important recall is given about the distinction between data and information, the former being the raw result of a measurement and the latter being an interpretation of data within a given applicative context, thus adding meaning to numbers. This clarification is important because data quality is a question that has aroused a lot of interest during the last decades, and numerous quality metrics have been defined and successfully used to control physical and numerical phenomenon. After a brief recall of the main data quality models and their underlying data metrics, MARCIN LENART emphases information quality and the different dimensions that may be envisaged to assess this quality (verifiability, reliability, relevance, etc.). Among existing models, the information quality model proposed by A. Revault d'Allonnes and M.-J. Lesot is chosen in this work. Motivation given for this choice is mainly the generic aspect of this trust-based model whose trust quantification relies on the aggregation of

four dimensions (reliability, competence, plausibility and credibility). MARCIN LENART then focuses on the specific case of information quality assessment when information is provided by sensors. Existing models to sensor quality measurement are presented. It is argued that existing approaches consider reliability as a main criterion whose quantification relies on a comparison between the sensor outputs and a ground truth. When several sensors are considered to control a process, then it is shown that the credibility dimension may play an important role in the final trust evaluation. Existing approaches dealing with credibility based on the sensor context are described. This part of the document allows MARCIN LENART to position his works and their objective, which is mainly to define a generic solution to information quality assessment that is thus independent from the type of sensor(s) considered, the availability of meta-knowledge (as e.g. ground truth) and that is able to observe how trust evolves according to time. *The distinction between data and information is clearly given, however the intersection between data quality and information quality is too quickly reviewed. A huge number of papers has been published on, possibly dynamic, data quality. It would have been interesting to have a deeper comparison between data quality and information quality. For instance, a topology of quality metrics and dimensions have been proposed by the W3C, and one can wonder where trust, as used in this PhD, is located in this topology. After having said that the information quality model proposed by A. Revault d'Allonnes and M.-J. Lesot is chosen in this work, it lacks a study and a discussion about the scientific and technological issues of adapting such a generic model to the considered specific case of sensors monitoring.*

The second part of the PhD is dedicated to the proposed ReCLiC model to information quality assessment. It is first recalled that the main motivation is to obtain a generic approach to the quality assessment of information provided by sensors. In the ReCLiC method the quality of the information provided by a sensor is quantified through a degree of trust that itself aggregates different quality dimensions, namely: reliability, competence, likelihood and credibility. To compute this degree of trust, ReCLiC relies only on the logs generated by sensors (sensor ID + message + timestamp + topic), a graph describing the expected sequences of logs, called a state transition graph, and a representation of the sensors network. It is said that this graph, which constitutes the only obstacle to a completely generic approach to sensor monitoring, may be inferred from the data or given a priori by a domain expert. It is then explained how the four dimensions used to quantify trust are interpreted in this context of sensor monitoring: reliability qualifies the sensor based on the number of errors generated, competence qualifies the sensor wrt. a topic, likelihood checks compatibility with previous messages based on a state transition graph, and credibility according to messages sent by other correlated sensors. It is then proposed to combine these four dimensions in a global trust degree that relies on a weighted mean that makes it possible to give more importance to internal (reliability, competence, likelihood) or external dimensions (credibility). *This part of the document explains in a clear and convincing way the differences between the manipulated*

quality dimensions and their complementarity. However, this section mainly describes an ad-hoc interpretation of an existing quality assessment model to the special case of event sensor monitoring. I was expecting MARCIN LENART *to get some distance by analysing the relevance of each dimension wrt. the problem addressed, to envisage other possible interpretations of the dimensions and to confront domain expert expectations with the knowledge each dimension provides. The question of the availability of a state transition graph and a network of sensors could also be further studied especially as the goal is to lead to a sensor monitoring approach as generic as possible. The four dimensions considered to analyse sensors capture very different aspects and the idea of combining them into a global trust degree should be justified. Is it more meaningful for a domain expert to have a global trust degree than to keep separated the four dimensions assessment?*

The third chapter is dedicated to the implementation of the ReCLiC model to a concrete application case about railway signalling with axle counters as sensors. After having clearly described this applicative context, an interpretation of the four quality dimensions involved in the ReCLiC model is proposed. Reliability (resp. competence) is seen as a ratio of error messages output by an axle counter within a time window (resp. wrt. a given topic), likelihood relies on an a priori defined state transition graph given by an expert or acquired from the MoTRicS2015 dataset, and credibility depends on messages sent by adjacent sensors. An analysis of the behaviour (limits and variations of the computed degree) of these four dimensions according to their parameters and the considered applicative context is then given. *This part of the document is interesting because the model is applied on a real applicative context. It shows how the ReCLiC model can be used. Some choices could be further justified especially by confronting expert knowledge and knowledge acquired from the data. For instance, could the expert state transition graph be completed or revisited using transitions observed in the dataset? Based on the given results, the question of keeping the four dimensions separated to better understand the observed phenomenon is worth to be asked.*

The fourth part gives the results of experimentations conducted on real data into which artificial errors have been added. Three types of error injections have been considered: uniform, burst (in a narrow time window) and random. The impact of these errors on the different quality dimensions and their aggregation as a trust degree have been studied in terms of degree decrease and speed of recovery. These experimentations confirm the analyses performed in the preceding chapter of the quality functions behaviour. Then, a study is provided to explain the impact of the ReCLiC parameters (weights and aggregation functions) on the obtained quality scores. *The conducted experimentations are very important to well understand the behaviour of the quality dimensions in the considered applicative context. They confirm that the expected behaviour is obtained in terms of alert (degree decrease) and history (speed of recovery). To justify that the way quality dimensions react according to received logs, it would have been interesting to have more explanations about the way domain experts would like*

to be informed when such situations occur. For instance, before having at their disposal the ReCLiC system, how experts interpret error messages, what are the actions taken, etc. This could help understand if the impact of messages on the quality degree and the speed of recovery to a normal situation are relevant or not.

The last part concerns experimentations done on real data (MoTRicS2015). MARCIN LENART also addresses in this part the question of how to render the domain expert the quality assessment provided by the ReCLiC model. This question is not easy especially due to the irregular production of logs. It is proposed to use heatmaps that are able to render both the temporal and geographical (adjacent sections) aspects of the generated logs. Two different situations, with various propagation effects on the adjacent sections, captured by ReCLiC and the proposed heatmap-based visualisation are described. *Heatmaps seem appropriate to render the quality evaluation performed by the ReCLiC model even if it could have been interesting to provide alternative and complementary visualisations, one for each dimension aggregated in the trust degree so as to help experts better understand why at a given time sensors are judged untrustworthy.*

Overall Evaluation The research works done and described by MARCIN LENART in this document address a very interesting issue. This problem is not easy, especially in the considered settings with hard constraints about the unavailability of additional meta-information and the objective of providing a unified approach to information quality assessment. The document is easy to read and to understand which shows the ability of MARCIN LENART to clearly describe a scientific problem and its application. The impact of this work, scientifically speaking, is nevertheless limited as it concerns the adaptation of an existing model to information quality assessment in a particular context of sensor monitoring. This impact could be enhanced by considering at least another application scenario so as to show the generic aspect of the ReCLic model, and by better describing the differences between domain expert expectations and knowledge generated by the proposed approach.

To conclude, I am convinced that the works done by MARCIN LENART during his PhD are worth to be defended.

Grégory Smits, the 25^{th} of November, 2019 (Lannion - France)