Review of "Mobile sensor network noise reduction and re-calibration using Bayesian network", by Xiang et al., 2015

Overview

New sensor technologies enable alternatives ways of measuring air quality, complementary to expensive (yet accurate) measurements by official monitoring stations. Small and low-cost sensors, such as the metal-oxide sensors used by the wearable M-Pods in this research, enable urban air pollution monitoring at a high spatio-temporal resolution. The current generation of these sensors, however, suffer from some serious issues which need to be addressed before they can be applied usefully. Sensor drift often becomes larger than the measured signal, which makes the readings useless unless timely recalibrated. Given the intended use of these devices by non-specialists, conventional recalibration is tedious and impractical. The paper presents a new method of re-calibrating in-field without requirements of ground truth, just by looking at the statistical properties of the data set produced by the sensors. The authors use a Bayesian network to correct for cross sensitivity of air pollutants and detect outliers. They extend the network with virtual evidence which enables them to compensate for sensor drift and update the sensor calibration functions iteratively.

General remarks

The presented work is innovative and is a promising method of solving low-cost air quality sensor issues. I recommend publication after major revision. The text can be written more concise, by restructuring and avoiding repetition. I miss a more clearly stated motivation of why this research is done, a more clearly stated overview of current sensor issues (Introduction). In the Conclusion section, I miss a more elaborate discussion of the results (applicability and limitations of the proposed method), and an outlook of future research (suggestions on how the method can be further improved).

It is mentioned that the M-Pods measure relative humidity, but these measurements are not used in the analysis of the results. To my knowledge, many low-cost sensors show a non-linear response to changes in RH. Did the researchers experience problems with this, and if yes, can the proposed method be used to compensate for it?

After noise removal, the M-Pod data stream is binned in one-minute intervals (Section 7.1.2). How does this time period relate to the response time of the different gas sensors to ambient concentrations? If the response time is longer than one minute, the time series will be strongly auto-correlated. Although this auto-correlation might be negligible for slowly varying signals (for instance during the side-by-side calibration period), it might become important for rapidly varying signals, for instance when the sensors are used in mobile applications. Does the proposed recalibration method suffer from auto-correlated signals?

The Bayesian network derives the conditional probability densities from a training period. How does the network react for new air pollution events (e.g. extreme winter or summer smog episodes) which were not included in the training set?

Throughout the paper it is mentioned that the error of the measurements increases when the sensor drifts. This is true of course, but these error are strongly biased. It would be better to make a clearer distinction between the error contribution due to bias and due to a random component.

Specific remarks

Abstract

To give the research a practical setting, I would mention here that the method was tested with metal oxide sensors measuring NO2, CO, and O3. Also consider to stress the innovative part of the research, namely extending the Bayesian network with virtual evidence for in-field sensor bias-correction/recalibration.

1. **Introduction**

Page 8972, line 20-22: “The traditional atmospheric researches (…) pollutant distributions.”
Unclear, rephrase.

Page 8973, line 11-12: “Drift is a phenomenon caused by many factors”.
What are typical time scales for drift to become significant?

Page 8973, line 15-16: “shifting the measurements results (…) without proper compensation”
Unclear

Page 8974, line 22: “The rest of this section”
“The rest of this paper”

2. **Motivation Example**

This section needs serious revision to be more informative and to avoid repetition. Consider moving sections 7.1.1 - 7.1.3 to this section.

3. **Related work**

Section 3.1
Sensor calibration should address at least two parameters: bias correction, due to drift (p_1 in equation (2)), and sensitivity correction, due to sensor degradation (p_2 in equation (2)). This is not very clear from this section.

Page 8976, line 1-2: “In contrast to the previous work, (…) various types of metal oxide sensors”
Is this because your method takes cross-sensitivity of the different sensors into account?

4. **System flow**

Section 4 and section 6.5 are describing the same system. Figure 1 and Figure 6 basically show the same idea. Avoid repetition.

5. **Basic Bayesian belief network**

Page 8978, line 18: “All the sensors are correlated”
Except from the temperature sensor.

Page 8978, line 18-21: “The readings (…) cross sensitivity”
This deserves a more prominent position, for instance in section 7.1.2

Page 8979, line 15-16: “followed by S, (…) followed by T”
T and *V* are used, depending on their context, for both Truth and Virtual, and Temperature and Voltage. Consider renaming *S*, *T*, *V* to e.g. *s*,*t*,*u* to avoid confusion.

6. Bayesian network with sensor re-calibration

Page 8980, line 14-15: “Significant drift (…) not re-calibrated.”
This can be stated more precisely. Sensor readings become useless if the bias is in the order of the measured signal.

Page 8980, line 17-27
This paragraph can be written more concise. It is sufficient to say that basic Bayesian networks are based on the assumption of unbiased measurements.

Page 8982, line 18: “virtue” → “virtual”

Page 8983, line 6: “The definition of the symbols can be found in sect. 5.2”
Put shortly in figure caption instead.

Equation (2)
Maybe I missed some conceptual ideas, but why is it not necessary to take cross sensitivity in the sensor functions into account, while it is included in the Bayesian network?

7. Experimental results

Section 7.1.1, 7.1.2, and 7.1.3 can be written more concise to avoid repetition. Maybe these section can be forwarded to Section 2, a it illustrates the research motivation.

Page 8988, line 4-7: “The red dots (…)blue dots”
This description could be moved to the caption of Figure 8.

Page 8988, line 15: “its error is increased more than 3 times”
Distinguish between bias and random error.

Page 8988, line 19-23: “It is demonstrated that drift is (…) using a predetermined model.”
This deserves a more prominent position, for example in the Conclusion section.

Table 2
Why not including the bias statistics for period 2, so we get a feeling of the evolution of the drift. Consider including statistics of the ground truth for comparison.

Page 8989, line 7-10: “Correlation percentage is defined (…) appropriate solution
It is unclear to me how the correlation percentage is defined.

Page 8989, line 25: “The program (…) 8GB memory”
This information is useless, unless reference is made to typical calculation times of the algorithm.

Page 8990, line 6: “is equal” → “is taken igual”

Page 8990, line 25: “3. Our technique”
It would be nice to give your method a name, or a short description (e.g. “Bayesian network with virtual evidence”), so that it can be easily referenced throughout this paper and in future publications.

Page 8991, line 17: ”Overall, (…) 34.1% on average”
Based on only three cases, this statement seems too general. Better would be: “In our setup, our technique can reduce error by 34% on average”. Also to be changed in Abstract and Conclusion.

Page 8991, line 26: “our technique is about 4 times better”
“our technique successfully recovers 4 times more data.”

Page 8992, line 16-25
Rewrite and clarify this paragraph.

8. Conclusion

See General remarks.

Page 8993, line 9-11: “Our method (…) recursively”
This statement would fit in nicely in the Abstract.