

機械学習を用いたセンサ値の補正手法の研究

大野雄一朗[†] 兼田千雅^{††} 井林宏文[‡] 峰野博史[‡]

[†]静岡大学情報学部 ^{††}静岡大学大学院総合科学技術研究科 [‡]静岡大学大学院情報学研究科

1 はじめに

温湿度を計測するためのセンサは多岐に渡る。中でもスイス Sensirion 社製デジタル温湿度センサは、湿度変化に対してセンサ素子電極間の静電容量の変化で計測する静電容量型の高精度デジタル湿度センサを搭載しており、湿度の測定範囲が 0~100%RH と広いだけでなく、湿度反応速度も数秒~10 秒程度と速いため、多様な計測機器や各種評価試験など幅広い分野で利用されている。その一種である SHT25 のデータシート [1]によると、相対湿度 10~90%RH までの湿度測定誤差は±2%RH、相対湿度 0%RH や 100%RH での湿度測定誤差は±4%RH、また温度 5~60℃までの温度測定誤差は±0.4℃、温度 0℃での温度測定誤差は±0.5℃というグラフが記載されている。一方、比較的正確に温湿度を測定できる乾湿計は、湿度計の基準器として空調業界等でも一般的に使用されている。それぞれ一長一短があり、デジタル温湿度センサは小型で手軽に導入可能であるが、アスマン通風乾湿計のように正確な温湿度計は小型とは言い難く、また湿球温度計のガーゼに水分を含ませ通風する必要があるなど、設置や運用の敷居が高い。

本研究では、計測条件によって計測値に誤差の生じやすいデジタル温湿度センサの計測値に対し、LOF (Local Outlier Factor) [2]と OC-SVM (One Class Support Vector Machine) [3]を用いた外れ値検出を行った後、機械学習を用いて高精度な計測値に補正する手法を提案する。

2 関連研究

関連研究として、安価なセンサを用いて空気品質データを高精度に補正する研究[4]や、汎用センサプラットフォームを用いて海底地形情報を補正によって取得する研究[5]が挙げられる。[4][5]ではフィルタリングによる平滑化が行われており、ノイズを除去できるという利点が存在する。一方で、平滑化ではノイズとともに必要なデータも消去してしまうという課題が存在する。また、必要なデータを消去せずにノイズのみ除去する方法として、外れ値検出がある。外れ値検出の手法として、全データを標準化し、標準化データと平均値の差の最大値が閾値を越えた場合、このデータを外れ値と判定する Smirnov-Grubbs 検定や、これを拡張し複数の外れ値を検出できる Tietjen-Moore 検

定などがある。しかし、これらの外れ値検出手法は 1 変数からの検出しか行えないため、温湿度のように季節などによって特徴が変化するデータの外れ値検出には不適であると言える。

3 補正手法

本研究では、高精度な計測値を取得するため機械学習を用いたセンサ値の補正を提案する。デジタル温湿度センサの計測値に対し外れ値検出を行った後、機械学習で温湿度センサデータをアスマン通風乾湿計のような高精度な計測値に補正する。

センサは電子の不規則な熱運動や電位差などから発生するノイズの影響を受けやすいため、高精度にデータを取得するためにはノイズの除去が必須である。ノイズの主な除去方法としてフィルタリングによる平滑化、外れ値検出が存在する。本研究では、必要なデータを消去する可能性が低い点、また複数の変数を扱える点から、外れ値検出手法である LOF と OC-SVM を用いる。LOF は近傍の密度からデータ p の外れ値の度合い LOF (p) (Local Outlier Factor of p) を求め、閾値を越えた場合は外れ値と判定する手法である。LOF (p) は他のデータがデータ p の近くにあるほど小さい値をとり、他のデータが遠くにあるほど大きい値をとるため、LOF (p) が大きいほど外れ値である可能性が高くなる。OC-SVM は SVM (Support Vector Machine) を拡張した ν -SVM[6]を使用し、全データを 2 クラスに分類する超平面を求めることで外れ値を検出する手法である。

LOF では外れ値判定に用いる閾値を、OC-SVM では外れ値とするデータの総数をそれぞれユーザが設定する必要がある。LOF では、LOF (p) の平均値の 5 倍を閾値とし、LOF (p) の値がこの閾値を超えたデータを外れ値であるとした。また、OC-SVM では LOF で定めた外れ値の

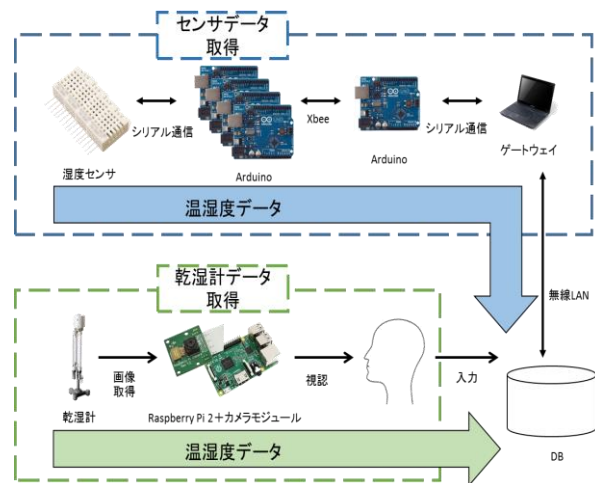


図 1 実験システム

Study on thermo-gyrometer sensing data correction based machine learning

Yuichiro Ohno[†], Yukimasa Kaneda^{††}, Hirofumi Ibayashi[‡], Hiroshi Mineno[†]

[†]Faculty of Informatics, Shizuoka University

^{††}National Institute of Information and Communications Technology, Shizuoka University

[‡]Graduation School of Informatics, Shizuoka University

表1 使用したデータ詳細

項目	温度	湿度
説明変数	温湿度センサ (温度)	温湿度センサ (温度, 湿度)
目的変数	乾湿計 (温度)	乾湿計 (湿度)
学習データ	2015/10/25 22:45～2015/10/27 16:05 (総データ数: 427 個)	
予測データ	2015/10/27 16:10～2015/10/28 23:55 (総データ数: 107 個)	
計測周期	5分	

数と同数のデータを外れ値であるとした。これらの検出手法に複数のセンサから取得した関連の強いデータを使用することで高精度な外れ値検出を目指す。

機械学習では、高い汎化能力を持ち未知データに強いとされる SVM を用いる。乾湿計と温湿度センサで温度、湿度を一定期間収集し、収集した温度、湿度データを学習データとしてモデルを作成する。作成したモデルで温湿度センサのデータから乾湿計の計測値を推定することで、温湿度センサが持つ運用容易性と、乾湿計が持つ高精度な計測値の取得を同時に実現する。

4 評価

機械学習を用いた補正手法の評価として予測実験を行った。評価指標として、外れ値検出を行わなかった場合と LOF, OC-SVM それぞれの外れ値検出を行った場合の RMSE (Root Mean Squared Error) を用い、RMSE の比較を行う。図1に実験システムを示す。乾湿計としてアスマン通風乾湿計 (SK-RHG-S) を使用し、Raspberry Pi 2 のカメラモジュールでアスマン通風乾湿計の目盛りを撮影した。その後、取得した画像から目視で計測値を取得した。また、デジタル温湿度センサとして SHT21, SHT25, DHT22 を用い、デジタル温湿度センサの計測値は Arduino を経由してデータベース (DB) に保存した。使用した学習データ、予測データの詳細を表1に示す。

予測実験の結果を図2に示す。湿度の予測誤差は、外れ値検出を行わなかった場合に比べ LOF では約 19%、OC-SVM では約 41% の誤差削減を確認した。しかし、温度においては大きな誤差削減が見られず、OC-SVM では約 8% の削減であり、LOF では約 5% 増加した。

予測実験では、外れ値検出手法に用いるデータとして表1の説明変数、目的変数を用いた。OC-SVM は SVM を拡張した ν -SVM を使用しているため、アルゴリズムが類似している。そのため、SVM, ν -SVM で行う各変数に対する重み付けも類似したものとなり、温度、湿度ともに外れ値検出手法として LOF を用いた場合と比べ予測誤差が小さくなったと考えられる。一方、LOF では各変数には重み付けをせず評価を行ったため、OC-SVM を用いた場合と比べ予測誤差が大きくなったと考えられる。

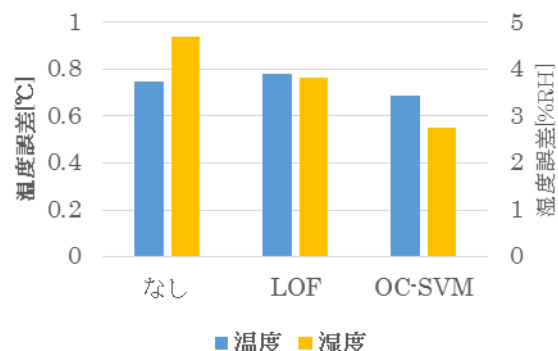


図2 予測実験結果

5 おわりに

本稿では幅広い分野で利用されている温湿度データをセンサのように手軽に、温湿度計のように高精度に取得するために、機械学習を用いたセンサ値の補正手法を提案し評価実験を行った。評価実験の結果、湿度において LOF では約 19%、OC-SVM では約 41% の誤差削減を確認した。一方、温度では LOF, OC-SVM ともに湿度ほどの誤差削減効果は見られなかった。今後は、温度での誤差削減を目指すとともに、SVM 以外での機械学習を用いた予測を行い、比較検証を行う予定である。

参考文献

- [1] SHT25
(https://www.sensirion.com/fileadmin/user_upload/customers/sensirion/Dokumente/Humidity_and_Temperature_Sensors/Sensirion_Humidity_and_Temperature_Sensors_SHT25_Datasheet_V3.pdf) (参照日 2015/12/25)
- [2] Breunig, Markus., et al., "LOF: identifying density-based local outliers", ACM sigmod record, Vol.29, No.2, pp.93-104, 2000.
- [3] Rätsch, Gunnar, et al., "SVM and boosting: One class", GMD-Forschungszentrum Informationstechnik, 2000.
- [4] Cheng, Yun, et al., "AirCloud: a cloud-based air-quality monitoring system for everyone", Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems, pp.251-265, 2014.
- [5] 畑中, 他, "水産業における情報技術の活用について-III. -Wavelet変換による高速ノイズ除去と海底地形図-", 日本航海学会論文集, Vol.118, pp.55-62, 2007.
- [6] Schölkopf, Bernhard, et al., "New support vector algorithms", Neural computation, Vol.12, Issue.5, pp.1207-1245, 2000.