

# 情報爆発時代におけるセンサデータ運用管理基盤の 実装と提案\*

山内 正人<sup>†</sup> 洞井 晋一<sup>‡</sup> 松浦 知史<sup>§</sup> 藤川 和利<sup>¶</sup> 砂原 秀樹<sup>||</sup>

奈良先端科学技術大学院大学\*\*

## 1 はじめに

広域センサネットワークでは、負荷分散を考慮すると共に、障害検知といったセンサデータの管理を行う必要がある。我々の提案する P2P ではマルチスケールでセンサデータを地理位置情報に基づいて分散管理することが可能である。そこで地理位置情報に基づいた分散環境上でデータの障害検知を行い、データに対して信頼度の付与を行う方式を提案した。この信頼度と地理位置情報を用いてデータ管理に最適なグルーピングを行い、グルーピングに基づいてデータをキャッシュしておくことで通信量を抑制すると共に、安定したデータの運用管理を可能としている。これを実際のセンサデータを用いて提案手法の有効性について評価を行った。

## 2 目的

広域センサネットワークではデータの信頼性が保障されていないためユーザがデータを信頼して利用することができない。センサデータの負荷分散手法については“Overlay Network Considering the Time and Location of Data Generation”[4]といった地理位置情報に基づいてデータを効率的に分散管理する研究が行われている。そこで本研究では地理位置情報に基づいた分散環境上でデータの信頼性確保を行う。

データの信頼性確保手法の事例として AMeDAS[1] での取り組みがある。AMeDAS は気象庁の地域気象観測システムである。この AMeDAS では取得したデータを 1 箇所に集め、障害検知を行い、設置されているセンサを目視による点検もしている。そのためユーザは信頼してデータを利用することが可能である (正常なデータの取得率 99.5%[3]) しかし、この手法を広域センサネットワークで適用した場合、大量のセンサを人手で信頼性確保することは困難である。また、負荷分散を考慮する必要があるため、1 箇所にデータを集めて障害検知を行うことも困難である。

## 3 運用管理基盤の提案

提案手法では分散環境上で障害検知を行い、障害検知結果を基に信頼度を付与する。また、付与された信頼度と地理位置情報を基にグルーピングを行う。地理位置に基づいてデータが分散管理しているため、信頼度が付与できればグルーピングを行うことが可能となる。このグルーピングは障害検知

やデータキャッシュに活用する。3.1 節で信頼度の説明を行い、3.2 節で障害検知手法の説明を行い、3.3 節でグルーピングの説明を行う。

### 3.1 信頼度

信頼度は障害検知結果を基にした障害回数の度合いである。障害可能回数の初期値は 0 で、この時信頼度は 0 となる。障害 (誤データ) が検知されなかった場合 1 (信頼度 10) づつ最大 10 (信頼度 100) まで増加する。ある時点で信頼度が 100 のセンサがあり、そのセンサにおいて障害が 1 回検知されると信頼度は 90 となる。しかし単一観測点で障害検知可能な手法 (後述) で障害と判断された場合は誤データの可能性が非常に高いため障害回数に関わらずデータの信頼度を 0 にする。信頼度を付与するためには信頼できる基準が必要となる。そこで本研究では AMeDAS のデータを基準とし常に信頼度 100 とする。

### 3.2 障害検知手法

気象データの障害検知手法として大きく分けて 3 つの手法がある。これらの手法によって障害検知を行い、結果を基に 3.1 節で述べた手順で信頼度を決定する。

- 単一観測点で障害検知可能な手法  
閾値との比較などによって障害検知を行う手法である。例えば過去最高気温 40.8 度といった値を閾値として設定しておく。センサデータがこの閾値を上回った場合誤データと判断する。
- 近隣データとの比較で障害検知可能な手法  
グルーピングされた近隣センサの値との差を設定値と比較することで障害検知を行う手法である。差が設定値を上回った場合誤データと判断する。
- 広域データとの比較で障害検知可能な手法  
広域でデータを比較することで局所的な気象状況発生時に有効な障害検知手法である。そのため降雨など局所的な気象要素が近隣データとの比較で誤データと判断された場合に適応し、誤障害検知を低減する。広域でデータを比較し近隣との差が設定値を上回る箇所が 1 箇所しかなかった場合誤データと判断する。

### 3.3 センサのグルーピング

基準となる AMeDAS のデータと全データを比較することができれば信頼度を決定することは容易である。しかし AMeDAS は約 20km 間隔でしか設置されておらず、距離が離れてしまうと誤差が大きくなり、誤検知の可能性が高くなってしまふ。そこで近隣の信頼度の高い 2 つのセンサとグループを形成し、比較することで信頼を確保する。1 つとの比較では比較相手に障害が発生していた場合に自センサが誤って障害と判断さ

\*Infrastructure for sensor data management and operation

<sup>†</sup>Masato Yamanouchi

<sup>‡</sup>Shinichi Doui

<sup>§</sup>Satoshi Matsuura

<sup>¶</sup>Kazutoshi Fujikawa

<sup>||</sup>Hideki Sunahara

\*\*Nara Institute of Science and Technology

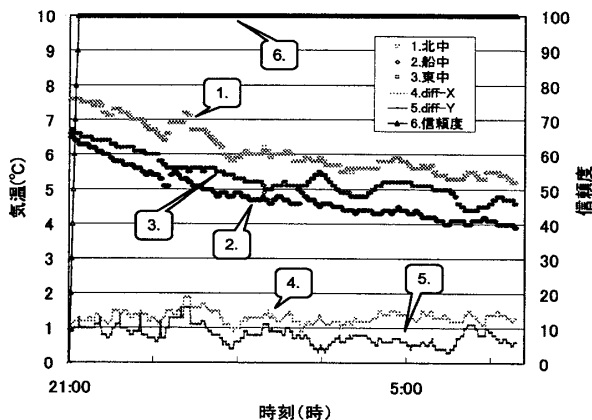


図 1: 正常時の信頼度の挙動

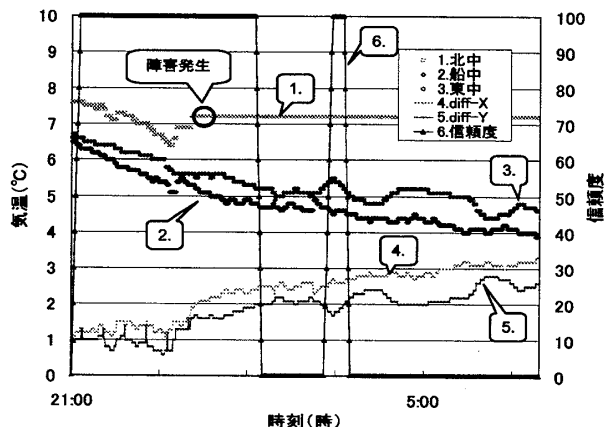


図 2: 異常時の信頼度の挙動

れてしまう可能性がある。しかし、比較相手が2つ同時に障害が発生する可能性は極めて低い。そのため2つと比較することで自センサが誤って障害と判断される可能性を軽減する。また、比較毎に比較相手が変化するとキャッシュしたデータが次の比較のときには利用できない。そこでグルーピングによって比較相手を半固定しておくことでキャッシュしたデータを利用することができ、トラフィック量の軽減や高速な障害検知が可能となる。

グルーピングの手法として、対象センサの近隣(例えば3km以内等を設定)にあるセンサデータを収集しセンサデータの信頼度が高い2つを選択する。センサデータが地理位置情報に基づいて分散管理されているため地理位置を指定した範囲検索がマルチスケールで容易に行える。同じ信頼度の場合にはあらかじめセンサ精度等を基に設定されている値によって高信頼度・高精度なセンサが選択されるようにする。AMeDASには最大値を設定し、高価で高精度なセンサほど高い値を設定する。それによって近隣にAMeDASがあればAMeDASと必ずグループを形成するようにする。

## 4 実験と考察

実験にはLive E! project[2]で2007年12月24日21時から25日8時に収集された実データを使用した。玉島北中学校(以下、北中)に設置されているセンサからのデータに対する信頼度の挙動について計測を行った。近隣比較で比較対象として利用した玉島東中学校(以下、東中)と北中との距離は約3km、船穂中学校(以下、船中)と北中との距離は約4kmである。また閾値として最大気温40.8度、最低気温-41度、過去1時間の気温変化量17度、近隣データとの誤差2度を使用した。計測結果を図1に示す。図中の凡例でdiff-Xは北中と船中の気温差、diff-Yは北中と東中の気温差、信頼度は北中の信頼度を示している。近隣誤差が2度以下で各閾値を下回っているため、信頼度は初期値0から増加していき最大値をたもっていることがわかる。次に午前0時から常に同じ値を示す異常が発生した場合を想定し、上記と同様の環境で計測を行った。計測結果を図2に示す。図中の凡例は図1に同じ。障害が発生してからデータの信頼度に反映されるまで1時間23分かかった。また、障害が発生しているが信頼度100または障害が発生していないが信頼度が100未満のデータ数を全体のデータ数で割った誤検知率は15%であった。内訳としては、障害が発生してから信頼度に反映されるまでが13%、反映された後の誤検知が2%となっている。

Live E!プロジェクトでは人手による障害検知を行っていたが、障害発生から検知にいたるまで3時間から数日かかっていた。数時間で検知した場合は広範囲のデータ集約を担当す

る計算機に障害が発生し、多数のセンサに障害が発生していた場合で単一のセンサにだけ障害が発生していた場合検知まで数日かかっている。そのため、単一のセンサの障害を1時間前後で検知できる本運用管理基盤は広域センサネットワークで有効であると考えられる。また、障害発生から信頼度に反映されるまでの時間及び誤検知率は気象状況によって前後すると考えられる。例えば急激な気温上昇や豪雨といった災害に繋がる可能性のある重要な気象変化のときに障害が発生した場合は短時間で設定値を越えるため反映時間が減少すると考えられる。障害が発生してから信頼度に反映されるまでの間における誤検知が誤検知全体の87%を占めているため、反映時間が減少することで誤検知率も大幅に軽減すると考えられる。

## 5 おわりに

本研究では、障害検知を行うために最適なグルーピングを行い、障害検知結果を基にセンサデータの信頼度を決定する運用管理基盤の提案を行った。広域センサネットワークではデータの信頼性が保障されていないためユーザがデータを信頼して利用することができない。そのためセンサデータに対して信頼度を付与することでユーザの要求にあったデータ提供を可能にする必要があった。そこで地理位置情報と信頼度から障害検知に最適なグルーピングを行い、障害検知結果を基にセンサデータの信頼度を決定することを行った。提案手法を実装し、Live E!プロジェクトで収集されている実データを用いて評価実験を行った。結果として、障害発生から1時間前後で信頼度に反映され、障害の誤検知率は15%であった。この検知時間や誤検知率は気象変化の度合いによって変化すると考えられるが、急激な気温上昇や豪雨といった重要な気象変化のときは実験結果より、高速・正確に検知可能であると考えられる。今後はその様な環境での評価実験を行い、その結果をもとにさらに高速・正確に検知可能にしていく。

## 参考文献

- [1] AMeDAS. <http://www.jma.go.jp/jp/amedas/>.
- [2] Live E! Project. <http://www.live-e.org/>.
- [3] 気象庁. <http://www.jma.go.jp/>.
- [4] Shinichi Doui, Satoshi Matsuura, Kazutoshi Fujikawa, and Hideki Sunahara. Overlay network considering the time and location of data generation. *IPJS Journal*, Vol. 49, No. 2, 2008.