

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6178218号
(P6178218)

(45) 発行日 平成29年8月9日(2017.8.9)

(24) 登録日 平成29年7月21日(2017.7.21)

(51) Int.Cl.

F 1

G06F 17/17 (2006.01)
G06F 17/18 (2006.01)G06F 17/17
G06F 17/18

Z

請求項の数 8 (全 15 頁)

(21) 出願番号 特願2013-236889 (P2013-236889)
 (22) 出願日 平成25年11月15日 (2013.11.15)
 (65) 公開番号 特開2015-97019 (P2015-97019A)
 (43) 公開日 平成27年5月21日 (2015.5.21)
 審査請求日 平成28年2月17日 (2016.2.17)

(73) 特許権者 000004226
 日本電信電話株式会社
 東京都千代田区大手町一丁目5番1号
 (74) 代理人 100119677
 弁理士 岡田 賢治
 (74) 代理人 100115794
 弁理士 今下 勝博
 (72) 発明者 倉沢 央
 東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内
 (72) 発明者 山本 淳
 東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内

最終頁に続く

(54) 【発明の名称】欠損値推定装置及び欠損値推定方法

(57) 【特許請求の範囲】

【請求項 1】

複数の変数を含むセンサデータのレコード群と当該レコード群における欠損値のうちの推定したい目的変数を受け付ける要求受付部と、

目的変数を有するレコードの変数のうち、欠損しておらず値を有する変数を第1の説明変数に設定し、前記レコード群のなかから前記第1の説明変数の揃ったレコードを解析対象レコード群として集約する集約部と、

前記集約部において集約された解析対象レコード群のレコードそれぞれについて、当該レコードと同一のオブジェクトから生成されたレコード群の前記第1の説明変数の統計値を算出して、当該レコードにおける前記第1の説明変数とは異なる第2の説明変数として追加する区間統計部と、

前記第2の説明変数が追加された解析対象レコード群について、前記第1及び第2の説明変数を用いた変数選択型の重回帰分析を行うことによって変数の係数の値を計算して当該値の絶対値が小さい変数を説明変数から除く変数選択型回帰分析部と、

重回帰分析の結果をもとに、目的変数を含むレコードの各説明変数の値及び係数から目的変数の値を計算する推定部と、

を備えることを特徴とする欠損値推定装置。

【請求項 2】

前記レコード群は、レコードを構成するセンサデータの観測時刻をレコードの生成時刻として含み、

前記区間統計部は、前記解析対象レコード群のうちの目的変数を有するレコードの生成時刻から一定期間内を設定範囲とし、当該設定範囲で生成された近傍レコード群について、同一のオブジェクトから生成されたレコード群の前記第1の説明変数の統計値を算出することを特徴とする請求項1に記載の欠損値推定装置。

【請求項3】

前記センサデータは、移動型センサノードにより観測されたデータであり、
前記レコード群は、レコードを構成するセンサデータの観測位置をレコードの生成位置として含み、

前記区間統計部は、前記解析対象レコード群のうちの目的変数を有するレコードの生成位置から一定距離内を設定範囲とし、当該設定範囲で生成された近傍レコード群について、同一のオブジェクトから生成されたレコード群の前記第1の説明変数の統計値を算出することを特徴とする請求項1又は2に記載の欠損値推定装置。 10

【請求項4】

前記区間統計部は、前記解析対象レコード群のレコードそれぞれについて、当該レコードと同一のオブジェクトから生成されたレコード群のうち、前記近傍レコード群に含まれるレコードの前記第1の説明変数の統計値を複数の異なる設定範囲の前記近傍レコード群について算出することを特徴とする請求項2又は3に記載の欠損値推定装置。

【請求項5】

複数の変数を含むセンサデータのレコード群と当該レコード群における欠損値のうちの推定したい目的変数を受け付けると、目的変数を有するレコードの変数のうち、欠損しておらず値を有する変数を第1の説明変数に設定し、前記レコード群のなかから前記第1の説明変数の揃ったレコードを解析対象レコード群として集約する集約手順と、

前記集約手順において集約された解析対象レコード群のレコードそれぞれについて、当該レコードと同一のオブジェクトから生成されたレコード群の前記第1の説明変数の統計値を算出して、当該レコードにおける前記第1の説明変数とは異なる第2の説明変数として追加する区間統計手順と、

前記第2の説明変数が追加された解析対象レコード群について、前記第1及び第2の説明変数を用いた変数選択型の重回帰分析を行うことによって変数の係数の値を計算して当該値の絶対値が小さい変数を説明変数から除く変数選択型回帰分析手順と、

重回帰分析の結果をもとに、目的変数を含むレコードの各説明変数の値及び係数から目的変数の値を計算する推定手順と、

を有することを特徴とする欠損値推定方法。

【請求項6】

前記レコード群は、レコードを構成するセンサデータの観測時刻をレコードの生成時刻として含み、

前記区間統計手順において、前記解析対象レコード群のうちの目的変数を有するレコードの生成時刻から一定期間内を設定範囲とし、当該設定範囲で生成された近傍レコード群について、同一のオブジェクトから生成されたレコード群の前記第1の説明変数の統計値を算出することを特徴とする請求項5に記載の欠損値推定方法。 30

【請求項7】

前記センサデータは、移動型センサノードにより観測されたデータであり、
前記レコード群は、レコードを構成するセンサデータの観測位置をレコードの生成位置として含み、

前記区間統計手順において、前記解析対象レコード群のうちの目的変数を有するレコードの生成位置から一定距離内を設定範囲とし、当該設定範囲で生成された近傍レコード群について、同一のオブジェクトから生成されたレコード群の前記第1の説明変数の統計値を算出することを特徴とする請求項5又は6に記載の欠損値推定方法。 40

【請求項8】

前記区間統計手順において、前記解析対象レコード群のレコードそれぞれについて、当該レコードと同一のオブジェクトから生成されたレコード群のうち、前記近傍レコード群

10

20

30

40

50

に含まれるレコードの前記第1の説明変数の統計値を複数の異なる設定範囲の前記近傍レコード群について算出することを特徴とする請求項6又は7に記載の欠損値推定方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、不完全センサデータの欠損値を推定する、欠損値推定装置及び欠損値推定方法に関する。

【背景技術】

【0002】

何らかの都合で観測できなかったり、観測した値を失ったりして、センサで観測された値がデータ化されたセンサデータの一部を欠損してしまうことはたびたび起こりうる。このような欠損を含むセンサデータを不完全センサデータとよぶ。例えば、センサの搭載された移動型センサノードにより観測されたセンシングデータを共有して活用する、ユーザ参加型のセンシング環境では、いつどこで観測されるかは各移動型センサノードの移動や行動に依存するため、すべての時空間のセンサ値を確実に得ることは難しい。さらに、移動型センサノードの仕様が一様でない場合は、すべての種類のセンサ値を得ることも難しい。それゆえ、ユーザ参加型のセンシング環境で集めたセンサデータは、時空間で整理すると、各移動型センサノードの不確実なセンシング行為に起因して、時空間やセンサの種類の欠損を含んでしまう。欠損値推定技術は、このようなユーザ参加型のセンシング環境において欠損してしまったセンサ値を推定する際に使われる。

【0003】

欠損値推定技術が対象とする不完全データの形式として、2つ以上の変数の値から構成されたレコードの群において、一部もしくはすべてのレコードでそれぞれ1つ以上の変数値が欠けているものがある。例えば、「時刻」や「緯度」、「経度」、「温度」、「湿度」、「降水量」、「紫外線」の7つの変数から構成されたレコード群において、「時刻」と「緯度」、「経度」の3つの変数値しか含まないレコードや、「時刻」と「温度」、「湿度」、「降水量」の4つの変数値を含むレコードなどが混在している、各レコードが必ずしもすべての変数値を含まないデータである。

【0004】

従来の欠損値推定技術は、欠損が生じた時刻や位置で観測できたセンサの値を説明変数、欠損したセンサの値を目的変数として、回帰推定していた（例えば、非特許文献1参照。）。つまり、欠損しているセンサと相関性の高いセンサを選別することに注力していて、装着方法や装着場所が多様なセンサノードから集めたデータを扱うことはそもそも想定されていなかった。回帰推定では、同じ時刻や位置で観測されたセンサの値から構成されるレコードを独立したものと見て、レコード群を学習データとしてそのまま利用していた。

【0005】

また、センサ端末を持つユーザの行動推定（例えば、座る、歩く、走るといった動作の推定）技術は、センサノードの装着場所（例えば、ズボンのポケット、胸ポケット、鞄の中）を推定することで、行動推定精度を向上していた（例えば、非特許文献2や非特許文献3参照。）。これらの技術では、装着場所を識別することを目的としていて、装着場所を特徴づける統計値を事前に選別しておき、推定時にはその選別結果を利用していた。具体的には、時間的に連続した加速度センサの観測値の統計値を特徴量として利用し、事前に学習した識別器（例えば、決定木やサポートベクターマシン）で行動をラベル付け（例えば、「歩く」、「走る」）していた。

【先行技術文献】

【非特許文献】

【0006】

【非特許文献1】倉沢央、佐藤浩史、山本淳、川崎仁史、中村元紀、松村一、「不完全センサデータの重回帰における変数選択手法」、信学技報、U S N 2 0 1 2 - 5 4、1 1 2

10

20

30

40

50

(242)、pp. 149 - 154、2012

【非特許文献2】倉沢央、川原圭博、森川博之、青山友紀、「センサ装着場所を考慮した3軸加速度センサを用いた姿勢推定手法」、情処研報2006-UBI-11-3、pp. 15-22、2006

【非特許文献3】藤波香織、「携帯機器の帶同場所のセンシング - このデータはどこから来たのか? - 」、情報処理、Vol. 54、No. 6、pp. 582 - 585、2013

【非特許文献4】H. Akaike, "A new look at the statistical model identification", IEEE Transactions on Automatic Control, Vol. 18, Issue 6, pp. 716 - 723, 1974

10

【非特許文献5】G. Schwarz, "Estimating the Dimension of a Model", Annals of Statistics, Vol. 6, No. 2, pp. 461 - 464, 1978

【発明の概要】

【発明が解決しようとする課題】

【0007】

複数センサ搭載の移動型センサノードで収集した不完全なセンサデータの欠損値を推定するにあたって、装着方法や装着場所が異なる移動型センサノードで観測されたセンサの値が混在しているとき、欠損値推定精度が低くなってしまう。例えば、装着方法に関しては、ズボンのポケットに装着した移動型センサノードと鞄の中に装着した移動型センサノードとでは、ズボンのポケットに装着した移動型センサノードの方が人の発汗の影響を受けて湿度が高いため、湿度センサの値が大きくなりやすくなる。また、装着場所に関しては、地面に近いところに装着された移動型センサノードと人の目の高さと同じくらいの位置に装着された移動型センサノードとでは、地面に近い方が日中のアスファルトの道路からの照り返しの影響が地面に近いほど大きくなるため、湿度センサの値が大きくなりやすくなる。このように、装着方法や装着場所によってセンサの値の傾向が変わってしまうため、移動型センサノードがどうやってどこに装着されているかを区別することが望ましい。

20

【0008】

従来の欠損値推定技術は、観測されたセンサの値のみを回帰推定に使っているため、時間的または空間的なセンサ値の変化によって特徴づけられるような装着方法や装着場所の区別が難しい。

30

【0009】

一方、従来の行動推定技術では、推定に用いるセンサの種類を限定したうえで、事前にユーザがセンサ端末を装着する方法や場所のすべての候補を学習して識別器を準備する必要があり、移動型センサノードのようにセンサの種類や、装着方法、装着場所の自由度が高く、ユーザが身につけるものとも限らないものに対してすべての装着方法や装着場所の候補に備えることは難しい。また、欠損を含むセンサデータでは、識別器で使う統計値を常に算出できるとも限らないため、そもそも欠損推定に適用することができない。

40

【0010】

さらに、行動推定技術では同一端末で時間的に連続したセンサの観測値の統計値は使っているが、同一端末で空間的に近いエリアで観測された値の統計値は用いていないため、地域的な統計値が装着方法や装着場所の特徴として表されるようなものへの対応が難しい。

【0011】

この発明は上記事情に着目してなされたもので、その目的とするところは、移動型センサノードの装着方法や装着場所を考慮して欠損値を推定できる欠損値推定装置及び欠損値推定方法を提供することにある。

【課題を解決するための手段】

【0012】

50

本発明は、センサデータの欠損値を精度高く推定するために、センサの値だけでなくこの統計値も説明変数として回帰推定に用いる。

【0013】

具体的には、本発明の欠損値推定装置は、

複数の変数を含むセンサデータのレコード群と当該レコード群における欠損値のうちの推定したい目的変数を受け付ける要求受付部と、

目的変数を有するレコードの変数のうち、欠損しておらず値を有する変数を第1の説明変数に設定し、前記レコード群のなかから前記第1の説明変数の揃ったレコードを解析対象レコード群として集約する集約部と、

前記集約部において集約された解析対象レコード群のレコードそれぞれについて、当該レコードと同一のオブジェクトから生成されたレコード群の前記第1の説明変数の統計値を算出して、当該レコードにおける前記第1の説明変数とは異なる第2の説明変数として追加する区間統計部と、

前記第2の説明変数が追加された解析対象レコード群について、前記第1及び第2の説明変数を用いた変数選択型の重回帰分析を行うことによって変数の係数の値を計算して当該値の絶対値が小さい変数を説明変数から除く変数選択型回帰分析部と、

重回帰分析の結果をもとに、目的変数を含むレコードの各説明変数の値及び係数から目的変数の値を計算する推定部と、

を備える。

【0014】

この構成によれば、統計値を説明変数に用い、相関性の低い変数を説明変数から除くため、センサデータの欠損値の推定精度を高めることができる。具体的な装着方法や装着場所を識別せずに、装着方法や装着場所を多少なりとも特徴づけうる統計値を説明変数として追加し、それらを変数選択型回帰分析で選別すれば、多様で曖昧な装着方法や装着場所を許容した欠損値推定ができる。つまり、欠損したセンサの値にほぼ同じ影響を及ぼす装着方法や装着場所は区別せずに扱うといったことができる。例えば、リュックサックに装着した場合と斜めがけバッグに装着した場合とで湿度センサの値への影響が区別できない程度の差だった場合、敢えて2つの装着方法を識別しないで欠損推定しても同じ推定精度が見込めるうえに、事前に識別のために「リュックサック」や「斜めがけバッグ」といった装着方法のラベルを準備しなくても良い。

【0015】

また、本発明の欠損値推定装置は、前記レコード群が、レコードを構成するセンサデータの観測時刻をレコードの生成時刻として含み、前記区間統計部は、前記解析対象レコード群のうちの目的変数を有するレコードの生成時刻から一定期間内を設定範囲とし、当該設定範囲で生成された近傍レコード群について、同一のオブジェクトから生成されたレコード群の前記第1の説明変数の統計値を算出してもよい。

この構成によれば、観測されたセンサの値だけでなく、移動型センサノードのセンサの値の時間的な変化をもとにした統計値を説明変数に加えることができる。これら統計値によって装着方法や装着場所を特徴づけることができる。例えば、ズボンのポケットに装着した移動型センサノードと鞄の中に装着した移動型センサノードとでは、加速度センサの値に注目すると、ある時刻における値の大きさで比較しても装着方法を区別するのが難しい。しかし、10秒間の値の変化の分散値で比較すると、ズボンのポケットに装着している場合の方が大きくなり、装着方法を特徴づける統計値としてこの分散値を利用できる。

【0016】

また、本発明の欠損値推定装置は、前記センサデータが、移動型センサノードにより観測されたデータであり、前記レコード群が、レコードを構成するセンサデータの観測位置をレコードの生成位置として含み、前記区間統計部は、前記解析対象レコード群のうちの目的変数を有するレコードの生成位置から一定距離内を設定範囲とし、当該設定範囲で生成された近傍レコード群について、同一のオブジェクトから生成されたレコード群の前記第1の説明変数の統計値を算出してもよい。

10

20

30

40

50

この構成によれば、移動型センサノードの装着方法や装着場所を特徴付ける統計量がセンサの値の空間的な変化をもとにしたものであっても対処できる。例えば、地面に近いところに装着された移動型センサノードと人の目の高さと同じくらいの位置に装着された移動型センサノードとでは、ダストセンサの値に注目すると、ある時刻における値の大きさや時間的な変化で比較しても装着場所を区別するのが難しい。砂利道やアスファルトといった路面環境によってダストが舞う度合いが変化しているような状況では、センサの値は時間よりも位置に依存して変化するためである。しかし、測定位置から10メートルの近傍で観測された値の平均値で比較すると、地面に近いところに装着している場合の方が大きくなり、装着場所を特徴づける統計値としてこの平均値を利用できる。

【0017】

10

また、本発明の欠損値推定装置は、前記区間統計部は、前記解析対象レコード群のレコードそれについて、当該レコードと同一のオブジェクトから生成されたレコード群のうち、前記近傍レコード群に含まれるレコードの前記第1の説明変数の統計値を複数の異なる設定範囲の前記近傍レコード群について算出してもよい。

この構成によれば、短期的なセンサの値の変化で表現できるものから長期的なセンサの値の変化で表現できるものまで、多様な装着方法や装着場所に対処することができる。

【0018】

具体的には、本発明の欠損値推定方法は、

複数の変数を含むセンサデータのレコード群と当該レコード群における欠損値のうちの推定したい目的変数を受け付けると、目的変数を有するレコードの変数のうち、欠損しておらず値を有する変数を第1の説明変数に設定し、前記レコード群のなかから前記第1の説明変数の揃ったレコードを解析対象レコード群として集約する集約手順と、

20

前記集約手順において集約された解析対象レコード群のレコードそれについて、当該レコードと同一のオブジェクトから生成されたレコード群の前記第1の説明変数の統計値を算出して、当該レコードにおける前記第1の説明変数とは異なる第2の説明変数として追加する区間統計手順と、

前記第2の説明変数が追加された解析対象レコード群について、前記第1及び第2の説明変数を用いた変数選択型の重回帰分析を行うことによって変数の係数の値を計算して当該値の絶対値が小さい変数を説明変数から除く変数選択型回帰分析手順と、

重回帰分析の結果をもとに、目的変数を含むレコードの各説明変数の値及び係数から目的変数の値を計算する推定手順と、

30

を有する。

【0019】

本発明は、統計値を説明変数に用い、相関性の低い変数を説明変数から除くため、センサデータの欠損値の推定精度を高めることができる。

【0020】

本発明の欠損値推定方法では、前記レコード群が、レコードを構成するセンサデータの観測時刻をレコードの生成時刻として含み、前記区間統計手順において、前記解析対象レコード群のうちの目的変数を有するレコードの生成時刻から一定期間内を設定範囲とし、当該設定範囲で生成された近傍レコード群について、同一のオブジェクトから生成されたレコード群の前記第1の説明変数の統計値を算出してもよい。

40

【0021】

本発明の欠損値推定方法では、前記センサデータが、移動型センサノードにより観測されたデータであり、前記レコード群が、レコードを構成するセンサデータの観測位置をレコードの生成位置として含み、前記区間統計手順において、前記解析対象レコード群のうちの目的変数を有するレコードの生成位置から一定距離内を設定範囲とし、当該設定範囲で生成された近傍レコード群について、同一のオブジェクトから生成されたレコード群の前記第1の説明変数の統計値を算出してもよい。

【0022】

本発明の欠損値推定方法では、前記区間統計手順において、前記解析対象レコード群の

50

レコードそれについて、当該レコードと同一のオブジェクトから生成されたレコード群のうち、前記近傍レコード群に含まれるレコードの前記第1の説明変数の統計値を複数の異なる設定範囲の前記近傍レコード群について算出してよい。

【発明の効果】

【0023】

本発明によれば、統計値を説明変数に加え、相関性の低い変数を説明変数から除くことで、搭載するセンサの種類や装着の仕方が不確実な移動型センサノードで集めたセンサデータから所望の欠損したセンサ値を推定する精度が向上する。これによれば、スマートフォンのような装着方法や装着場所をある程度限定できるデバイスだけでなく、人や動物の体や車などのどこにでも貼り付けられる小型センサノードを使って多様な状況で集めたセンサデータからの欠損値推定を本発明では実現できる。

10

【図面の簡単な説明】

【0024】

【図1】本実施形態に係る欠損値推定装置の一例を示す構成図である。

【図2】オリジナルレコード群の一例を示す。

【図3】欠損値推定方法の一例を示すフローチャートである。

【図4】解析対象レコード群の一例を示す。

【図5】端末ID = 1についての直近5, 10分のレコードの一例を示す。

【図6】端末ID = 1についての近傍100mのレコードの一例を示す。

20

【図7】端末ID = 1についての直近5, 10分の統計値の一例を示す。

【図8】端末ID = 1についての近傍100mの統計値の一例を示す。

【図9】端末ID = 3についての直近5, 10分のレコードの一例を示す。

【図10】端末ID = 3についての近傍100mのレコードの一例を示す。

【図11】端末ID = 3についての直近5, 10分の統計値の一例を示す。

【図12】端末ID = 3についての近傍100mの統計値の一例を示す。

【図13】統計値を追加した解析対象レコード群の一例を示す。

【発明を実施するための形態】

【0025】

以下、本発明の実施形態について、図面を参照しながら詳細に説明する。なお、本発明は、以下に示す実施形態に限定されるものではない。これらの実施の例は例示に過ぎず、本発明は当業者の知識に基づいて種々の変更、改良を施した形態で実施することができる。なお、本明細書及び図面において符号が同じ構成要素は、相互に同一のものを示すものとする。

30

【0026】

本発明の欠損値推定方法は、レコードごとに観測時刻から一定期間内に同一の移動型センサノードで観測されたセンサの値、または、観測位置から一定距離内に同一の移動型センサノードで観測されたセンサの値をもとに装着方法や装着場所を表す統計値を、レコードごとに算出する。そして、欠損が生じた時刻や位置で観測できたセンサの値だけでなくこの統計値も説明変数として変数選択型回帰で推定に使い、欠損値を推定する。

40

【0027】

図1に本発明の欠損値推定装置10の構成を示す。欠損値推定装置は、要求受付部13、集約部14、区間統計部15、変数選択型回帰分析部16、推定部17、結果応答部18の各処理部と、第1メモリ11、第2メモリ12の各記憶部を備える。

【0028】

図2に本発明の欠損値推定装置に入力されるレコードの例を示す。各レコードは、レコードID、端末ID、時刻、緯度、経度、ダスト、加速度、温度、湿度、騒音の変数値から構成されている。「NULL」は欠損を意味し、「?」は欠損かつ推定したい値を意味する。移動型センサノードの装着方法や装着場所はレコードに記録されていないが、端末IDのような個々の移動型センサノードを区別する変数が記録されているものとする。例えば、IDが13のレコードは、移動型センサノードの端末IDが1、観測時刻が13:2

50

6、緯度 35.4145、経度 139.4516、ダスト 9.40、加速度 6.01、湿度 37.73、騒音 62.04 で、温度の値が欠損かつ推定したい値であることを示している。この場合、温度の値が目的変数となる。

【0029】

以降では、欠損値推定の要求を受け付けて推定値を応答するまでを、図 3 のフローチャートと各処理部や各記憶部を参照しながら説明する。本実施形態に係る欠損値推定方法は、集約手順と、区間統計手順と、変数選択型回帰分析手順と、推定手順と、を順に有する。集約手順では、ステップ S 1 ~ S 3 を実行する。区間統計手順では、ステップ S 4 を実行する。変数選択型回帰分析手順では、ステップ S 5 ~ S 6 を実行する。推定手順では、ステップ S 7 ~ S 8 を実行する。本発明の特徴である装着方法や装着場所の区別のための手順はステップ S 4 で実行する。

10

【0030】

ステップ S 1 について説明する。要求受付部 13 は、欠損を含むレコード群と推定したい値の要求を受ける。これをオリジナルレコード群として第 1 メモリ 11 に書き込む。

【0031】

ステップ S 2 について説明する。集約部 14 は、推定対象である変数の値を求めるべきレコードを「クエリ」、推定したい変数を「目的変数」、欠損していない変数を「説明変数」とする。図 2 の例では、クエリは ID が 13 のレコード、目的変数は温度、説明変数は時刻、緯度、経度、ダスト、加速度、湿度、騒音となる。レコード ID = 13 のうちの「速度」が「NUL」である場合は、「速度」は説明変数から除かれる。

20

【0032】

ステップ S 3 について説明する。集約部 14 は、第 1 メモリ 11 のオリジナルレコード群のうち、目的変数と説明変数が揃ったレコードを集約し、解析対象レコード群として第 1 メモリ 11 に書き込む。解析対象レコード群は、学習データである。図 4 に解析対象レコード群の例を示す。

【0033】

本実施形態では、図 4 の例では、クエリの ID = 13 のレコードに加えて、目的変数の温度センサの値と説明変数の時刻、緯度、経度、ダスト、加速度、温度、湿度、騒音の値が「NUL」ではないレコードが集約される。例えば、図 2 に示すレコード ID が 1 から 12 は温度の値が「NUL」となっているため、解析対象レコード群からは除外される。また、レコード ID が 14 から 16 はダストと湿度の値が「NUL」となっているため、同様に除外される。図 2 に示すレコード ID = 27, 28, 805, 3109 については目的変数と説明変数が揃っているため、解析対象レコードとなる。

30

【0034】

ステップ S 4 について説明する。区間統計部 15 は、解析対象レコード群のうちの同一のオブジェクトから生成されたレコード群ごとに統計値を算出し、各レコードの説明変数に追加する。同一のオブジェクトから生成されたレコード群は、例えば端末 ID を有するレコード群である。統計値は、端末 ID、時刻、位置（緯度と経度）及び目的変数を除いた 1 つ又は複数種類のセンサの値を用いるものとする。このとき、期間又は距離の近傍レコード群を集約し、集約した近傍レコードについてのみ統計値を算出することが好ましい。

40

【0035】

本実施形態では、レコード ID = 13 のレコードは端末 ID = 1 であるため、端末 ID = 1 の期間又は距離の近傍レコードを集約する。期間の近傍レコードは、例えば、時刻の値が、レコード ID = 13 の時刻「13:26」から一定期間内のレコードである。ここでの一定時間は、5 分以内、10 分以内のように、指標間隔で集約する。距離の近傍レコードは、例えば、位置の値が、レコード ID = 13 の緯度「35.4145」、経度「139.4516」で特定される位置から一定距離内に位置するレコードである。ここでの一定距離は、100 m 以内のように、指標間隔で集約する。

【0036】

50

期間については、図5に示すように、レコードID = 13の時刻から直近5分以内に測定されたレコードID = 11, 12, 13と直近10分以内に測定されたレコードID = 8, 9, 10の2区画を集約する。距離については、図6に示すように、レコードID = 13の観測地から100m以内で観測されたレコードID = 7~13の1区画を集約する。また統計値は分散値と平均値とする。

【0037】

図7は、図5に示す期間について集約したレコード群の統計値である。図8は、図6に示す距離について集約したレコード群の統計値である。図7及び図8に示すレコードが、レコードID = 13のレコードに説明変数として追加される。

【0038】

ステップS4では、端末ID = 1と同様に、解析対象レコード群に含まれる端末ID = 3, 28, 227についても、期間又は距離の近傍レコード群を集約し、統計値を算出する。図9に、端末ID = 3についての期間の近傍レコード群の一例を示す。図10に、端末ID = 3についての距離の近傍レコード群の一例を示す。図11に、図9に示す期間について集約したレコード群の統計値を示す。図12に、図10に示す距離について集約したレコード群の統計値を示す。

【0039】

ここで、例えば、カバンに装着されたときに、直近5分と直近10分の加速度の分散値といったように区切る幅を変えることで、カバンの中を探るときの振動のような詳細な特徴と、カバンを持って歩きまわるときの振動のような大局的な特徴の両方を後の変数選択型回帰分析のときに選別対象にできる。

【0040】

なお、本発明のいう統計値は上記特徴を満たす様々な算出方法がその対象になるものであって、分散値や平均値に限られない。具体的に一例を挙げると、最大値や最小値、中間値、周波数スペクトルのような統計値であっても良い。

【0041】

図13に、ステップS4で統計値を追加したレコード群の一例を示す。図13では、解析対象レコード群の各レコードについて、同様に近傍レコード群の分散値や平均値を算出した結果の例を示す。後のステップの説明のため、端末IDが重複しているものはこの例では省いて示している。また、後のステップの説明のため、端末ID = 1は「車」に、端末ID = 3は「バギー」に、端末ID = 28, 54, 58は「斜めがけバッグ」に、端末ID = 97, 183, 227は「ハンドバッグ」に、端末ID = 312は「人のズボンのポケット」に装着したものとする。

【0042】

ステップS5について説明する。変数選択型回帰分析部16は、変数選択型の重回帰分析をして、結果を第2メモリ12に書き込む。変数選択型の重回帰分析は以下の式で算出される。

【数1】

$$\min_{\theta} \sum_i \left\| W(R_i) \left(\frac{R_i(q) - \mu(q)}{\sigma(q)} - \sum_{v \in R_i} \theta(v) \frac{R_i(v) - \mu(v)}{\sigma(v)} \right) \right\|^2 + \lambda \|\theta\|^l$$

ただし、Rは対象レコード群、R_iはレコードIDがiのレコード、R_i(v)はR_iの変数vの値、W(R_i)はR_iの重み、qは目的変数、θは説明変数の係数、(v)は変数vの係数の値、μ(v)は変数vの値の平均値、σ(v)は変数vの値の標準偏差、LはL1正則化のパラメータを表す。

【0043】

なお、本発明のいう変数選択型の重回帰分析は、様々な変数選択方法がその対象になるものであって、L1正則化による計算に限られない。具体的に一例を挙げると、変数ごとに回帰係数を計算して係数の絶対値の小さいものを削減する方法や、Akaike Information Criterion (AIC) (例えば、非特許文献4参照。) や

10

20

30

40

50

Schwarz's Bayesian Information Criterion (BIC) (例えば、非特許文献5参照。)といったモデル選択基準をもとにしたものであってもよい。さらに、欠損が多く含まれる場合は、欠損推定に用いる変数を絞り込みつつ解析対象レコード群を更新していく変数選択型重回帰分析(例えば、非特許文献1参照)であってもよい。

【0044】

ステップS6について説明する。変数選択型回帰分析部16は、重回帰分析の結果をもとに、相関性の低い変数を説明変数から除く。ここでいう相関性の低い変数とは、ステップS5で算出された変数の係数がほぼ0の値となるものを指す。相関性の低い変数は、例えば、0.01以下とする。本実施形態では、時刻、直近10分の加速度の分散値、近傍100mの騒音の平均値が、相関の高い説明変数として選ばれ、他の説明変数は除かれるとする。

【0045】

ステップS7について説明する。推定部17は、重回帰分析の結果をもとに、クエリの各変数から推定したい値を計算する。目的変数qの推定値Q(q)は(数式2)にもとづいて算出される。

【数2】

$$Q(q) = \sum_{v \in n} \theta(v) Q(v)$$

ただし、(v)は変数vの係数の値、Q(v)は変数vの値である。

【0046】

本実施形態では、ステップS5における変数選択型回帰分析部16によって各説明変数の係数の値が算出され、当該係数に基づいて相関の低い説明変数が除かれている。本ステップで残されている説明変数は、時刻、直近10分の加速度の分散値、近傍100mの騒音の平均値である。そして、各説明変数の係数が、0.2、0.3、0.1の場合、推定部17において算出される推定式は、(数式2)より次式のようになる。

(数3)

(温度) = 0.2 * (時刻) + 0.3 * (直近10分の加速度の分散値) + 0.1 * (近傍100mの騒音の平均値) + (定数)

【0047】

ステップS8について説明する。結果応答部18はステップS7で計算された推定値を出力する。

【0048】

以上の手順で、欠損値推定装置10は不完全データの欠損した値を推定することができる。

【0049】

以上説明したように、本発明は、図2に示すオリジナルレコード群に含まれているそのままの変数を説明変数に用いて重回帰分析で求めたとしても、各変数の相関性は低い。これに対して、本実施形態に係る発明では、説明変数にダスト、加速度、湿度、騒音の統計値を加えている。本実施形態に係る発明は、これらの統計値を用いて、目的変数とより相関性の高い変数である、直近10分の加速度の分散値と近傍100mの騒音の平均値の2つの説明変数を温度センサの値の推定式に加えることができた。これにより、本実施形態に係る発明は、センサデータの欠損値を精度高く推定することができる。

【0050】

従来の行動推定技術で本実施形態の課題に取り組んだ場合は、事前に「車」、「斜めがけバッグ」、「ハンドバッグ」、「人の胸ポケット」、そして「人のズボンのポケット」といった装着方法や装着場所の候補と識別に有効な統計値を決めておき、直近の連続したセンサの値の統計値から、それら5つの候補に識別することになる。例えば、近傍100mの騒音の平均値が大きいものは「車」に識別する、といった処理であった。これに対して、本実施形態に係る発明では、直近10分の加速度の分散値と近傍100mの騒音の平

10

20

30

40

50

均値の2つの統計値を説明変数として欠損推定にそのまま利用して、装着方法や装着場所が欠損したセンサ値に与える影響を反映させた推定式が得られた。本実施形態では、「車」のような騒音の平均値の大きいものが温度を高くする傾向や、「ズボンのポケット」のような加速度の分散値の大きいものが同様に温度を高くする傾向を推定式に反映できた。また、「斜めがけバッグ」と「ハンドバッグ」のように温度の値に与える影響がほぼ同じようなものは区別なく扱えた。

【 0 0 5 1 】

なお、本発明の装置はコンピュータとプログラムによっても実現でき、プログラムを記録媒体に記録することも、ネットワークを通して提供することも可能である。

【産業上の利用可能性】

10

(0 0 5 2)

本発明は情報通信産業に適用することができる。

【符号の説明】

[0 0 5 3]

1.0 · 欠損値予測装置

11 · 第1卷

12 · 第2章

13：要求受

14：集約部

1.5. 区間統計部

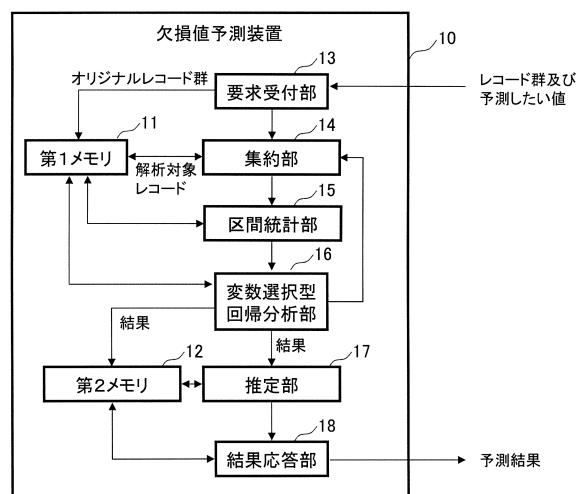
15：区间統

16：変数選

1 6 : 变数选择型

30

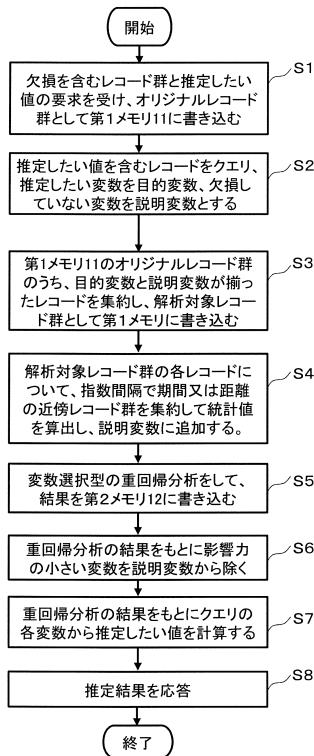
〔 図 1 〕



【图2】

レコードID	端末ID	時刻	緯度	経度	ダスト	加速度	温度	湿度	騒音
1	1	13:02	35.41.39	139.45.06	8.86	10.04	NUL	38.29	57.59
2	1	13:04	35.41.40	139.45.07	10.28	4.94	NUL	37.78	68.51
3	1	13:06	35.41.40	139.45.08	8.76	14.87	NUL	37.73	57.29
⋮									
10	1	13:20	35.41.38	139.45.12	10.90	12.59	NUL	38.16	58.18
11	1	13:22	35.41.40	139.45.14	11.57	13.24	NUL	38.24	49.90
12	1	13:24	35.41.42	139.45.15	11.13	10.43	NUL	37.50	52.09
13	1	13:26	35.41.45	139.45.16	9.40	6.01	?	37.73	62.04
14	2	13:02	35.48.36	139.41.86	NUL	9.97	23.56	NUL	54.38
15	2	13:04	35.48.36	139.41.87	NUL	9.62	23.44	NUL	62.33
16	2	13:06	35.48.38	139.41.89	NUL	9.90	23.42	NUL	50.38
⋮									
27	3	13:02	35.48.42	139.42.10	13.13	9.75	23.36	41.81	55.36
28	3	13:04	35.48.42	139.42.12	12.02	9.86	23.32	41.72	52.04
⋮									
805	28	13:35	35.40.19	139.42.59	10.07	10.4	22.2	36.6	61.0
⋮									
3109	227	14:04	35.41.38	139.42.92	8.6	8.1	21.9	30.1	59.0

【図3】



④ で

【図4】

解析対象レコード群						
レコードID	端末ID	時刻	緯度	経度	ダスト	加速度
1	1	13:02	35.4139	139.4510	8.86	10.04
13	1	13:26	35.4145	139.4516	9.40	6.01
27	3	13:02	35.4142	139.4512	12.02	9.86
28	3	13:04	35.4142	139.4512	12.02	9.86
805	28	13:35	35.4019	139.4259	10.7	10.4
3109	227	14:04	35.4138	139.4292	8.6	8.1
					21.9	30.1
					59.0	

【図5】

端末ID=1についての直近5.10分のレコード						
レコードID	端末ID	時刻	緯度	経度	ダスト	加速度
8	1	13:16	35.4140	139.4510	10.39	8.86
9	1	13:18	35.4139	139.4510	9.49	9.94
10	1	13:20	35.4138	139.4512	10.90	12.59
11	1	13:22	35.4140	139.4514	11.57	13.24
12	1	13:24	35.4142	139.4515	11.13	10.43
13	1	13:26	35.4145	139.4516	9.40	6.01
					?	37.73
					62.04	

端末ID=1についての近傍100mのレコード

レコードID	端末ID	時刻	緯度	経度	ダスト	加速度
7	1	13:14	35.4139	139.4510	9.87	9.56
8	1	13:16	35.4140	139.4510	10.39	8.86
9	1	13:18	35.4139	139.4510	9.49	9.94
10	1	13:20	35.4138	139.4512	10.90	12.59
11	1	13:22	35.4140	139.4514	11.57	13.24
12	1	13:24	35.4142	139.4515	11.13	10.43
13	1	13:26	35.4145	139.4516	9.40	6.01
					?	37.73
					62.04	

【図6】

【図7】

端末ID=1についての直近5、10分の統計値

レコードID	端末ID	時刻	緯度	経度	直近5分の分散値			直近10分の分散値			近傍10分の平均値			
					ダスト	加速度	湿度	騒音	ダスト	加速度	湿度	騒音	ダスト	加速度
27	3	13:02	35.4839	139.4210	13.13	9.75	23.36	41.81	53.36	52.04	55.61	59.39	55.5	55.5
28	3	13:04	35.4842	139.4212	12.02	9.86	23.32	41.72	52.04	52.04	55.61	59.39	55.5	55.5
29	3	13:06	35.4843	139.4213	13.14	9.79	23.30	41.89	53.36	52.04	55.61	59.39	55.5	55.5
30	3	13:08	35.4844	139.4215	11.75	9.76	23.49	41.43	52.04	52.04	55.61	59.39	55.5	55.5
31	3	13:10	35.4846	139.4217	11.84	10.02	23.48	42.58	52.04	52.04	55.61	59.39	55.5	55.5
32	3	13:12	35.4846	139.4217	11.82	9.79	23.69	42.02	52.04	52.04	55.61	59.39	55.5	55.5

【図9】

端末ID=3についての直近5、10分の統計値

レコードID	端末ID	時刻	緯度	経度	直近5分の分散値			直近10分の分散値			近傍100mの平均値			
					ダスト	加速度	湿度	騒音	ダスト	加速度	湿度	騒音	ダスト	加速度
28	3	13:04	35.4842	139.4212	12.02	9.86	23.32	41.72	52.04	52.04	55.61	59.39	55.5	55.5
29	3	13:06	35.4843	139.4213	13.14	9.79	23.30	41.89	53.36	52.04	55.61	59.39	55.5	55.5
30	3	13:08	35.4844	139.4215	11.75	9.76	23.49	41.43	52.04	52.04	55.61	59.39	55.5	55.5
31	3	13:10	35.4846	139.4217	11.84	10.02	23.48	42.58	52.04	52.04	55.61	59.39	55.5	55.5
32	3	13:12	35.4846	139.4217	11.82	9.79	23.69	42.02	52.04	52.04	55.61	59.39	55.5	55.5

【図10】

端末ID=3についての近傍100mの統計値

レコードID	端末ID	時刻	緯度	経度	直近5分の分散値			直近10分の分散値			近傍100mの平均値			
					ダスト	加速度	湿度	騒音	ダスト	加速度	湿度	騒音	ダスト	加速度
28	3	13:04	35.4842	139.4212	12.02	9.86	23.32	41.72	52.04	52.04	55.61	59.39	55.5	55.5
29	3	13:06	35.4843	139.4213	13.14	9.79	23.30	41.89	53.36	52.04	55.61	59.39	55.5	55.5
30	3	13:08	35.4844	139.4215	11.75	9.76	23.49	41.43	52.04	52.04	55.61	59.39	55.5	55.5
31	3	13:10	35.4846	139.4217	11.84	10.02	23.48	42.58	52.04	52.04	55.61	59.39	55.5	55.5
32	3	13:12	35.4846	139.4217	11.82	9.79	23.69	42.02	52.04	52.04	55.61	59.39	55.5	55.5

【図8】

【図 1 1】

レコード番号	ID	時刻	緯度	経度	直近10分の分散値		直近10分の平均値		近傍100mの分散値		近傍100mの平均値		近傍10分の分散値		近傍10分の平均値																												
					ダスト	加速度	ダスト	加速度	ダスト	加速度	ダスト	加速度	ダスト	加速度	ダスト	加速度																											
13	1	13:23:55.41	139.4°S	60°E	?	31.7	62.0	1.5	0.8	0.0	49.5	0.8	6.9	0.1	19.2	0.3	8.2	31.6	57.1	0.5	10.2	31.9	55.5	0.7	5.8	0.1	28.4	10.4	10.1	37.8	56.8												
32	2	13:13:55.48	139.4°S	61.8°E	9:39	236	42.0	59.6	0.0	0.0	0.3	1.3	0.4	0.0	0.1	123.7	11.9	9.9	42.0	60.2	12.3	9.8	41.9	57.3	0.0	0.0	0.3	1.3	0.4	0.0	0.1	12.3	11.8	9.9	42.0	60.2	12.3	9.8	41.9	57.3			
805	28	13:3:35.40	139.4°S	61.9°E	2:17	2	6	0.0	0.0	0.3	1.3	0.4	0.0	0.1	123.7	11.9	9.9	42.0	60.2	12.3	9.8	41.9	57.3	0.0	0.0	0.2	14.3	12.1	11.8	41.9	57.3	0.0	0.0	0.2	14.3	12.1	11.8	41.9	57.3				
1804	54	13:3:35.40	139.4°S	61.9°E	7:49	7	9	22.3	23.2	51.1	1.3	9.4	0.0	0.0	50.1	2.1	9.1	0.1	200.1	11.2	10.6	35.7	62.0	1.9	10.6	37.1	62.5	2.3	9.3	0.3	20.4	10.0	9.3	37.3	52.2	0.0	0.0	0.2	14.3	12.1	11.8	41.9	57.3
2280	58	13:4:35.41	139.4°S	61.9°E	10:0	10.0	22.3	26.9	61.7	1.3	9.7	0.0	0.0	47.3	2.1	9.5	0.1	18.9	0.0	10.6	29.3	62.0	10.3	0.6	23.5	62.1	2.1	9.6	0.8	19.4	10.1	9.6	42.6	52.3	0.0	0.0	0.2	14.3	12.1	11.8	41.9	57.3	
2593	97	13:5:35.41	139.4°S	62.0°E	0:07	9	3	22.3	28.7	53.6	1.4	10.6	0.0	0.0	49.6	2.2	9.1	0.4	19.8	1.0	9.7	30.3	54.0	1.5	0.5	30.8	54.2	3.2	9.9	1.2	20.4	10.0	9.3	40.1	52.6	0.0	0.0	0.2	14.3	12.1	11.8	41.9	57.3
2873	183	14:0:35.40	139.4°S	62.0°E	9:27	9	9	22.1	34.0	61.8	1.5	10.1	0.0	0.0	59.0	3.0	9.4	1.0	20.0	0.5	10.0	34.3	62.3	10.4	1.0	34.9	62.5	3.5	9.7	1.4	20.3	9.9	10.0	39.4	50.0	0.0	0.0	0.2	14.3	12.1	11.8	41.9	57.3
3109	287	14:0:35.41	139.4°S	62.0°E	8:31	21.9	36.1	59.0	1.3	9.8	0.0	48.7	1.7	8.7	0.3	19.5	3.1	9.0	30.9	59.6	9.2	8.7	31.1	60.2	1.8	9.0	0.4	19.8	6.5	10.3	39.9	50.0	0.0	0.0	0.2	14.3	12.1	11.8	41.9	57.3			
3268	312	14:0:35.42	139.4°S	62.0°E	17:18	14.2	9.8	22.7	28.9	60.3	2.3	9.3	0.0	0.0	50.2	2.0	10.4	0.1	20.1	15.0	0.7	30.6	61.3	15.7	11.7	30.9	62.0	2.1	10.9	0.7	20.7	10.2	10.0	39.8	53.8	0.0	0.0	0.2	14.3	12.1	11.8	41.9	57.3

統計値を追加したレコード群

【図 1 3】

レコード番号	ID	時刻	緯度	経度	直近10分の分散値		直近10分の平均値		近傍100mの分散値		近傍100mの平均値		直近10分の分散値		直近10分の平均値																												
					ダスト	加速度	ダスト	加速度	ダスト	加速度	ダスト	加速度	ダスト	加速度	ダスト	加速度																											
2593	97	13:5:35.41	139.4°S	62.0°E	0:07	9	3	22.3	28.7	53.6	1.4	10.6	0.0	0.0	49.6	2.2	9.1	0.4	19.8	1.0	9.7	30.3	54.0	1.5	0.5	30.8	54.2	3.2	9.9	1.2	20.4	10.0	9.3	40.1	52.6	0.0	0.0	0.2	14.3	12.1	11.8	41.9	57.3
2873	183	14:0:35.40	139.4°S	62.0°E	9:27	9	9	22.1	34.0	61.8	1.5	10.1	0.0	0.0	59.0	3.0	9.4	1.0	20.0	0.5	10.0	34.3	62.3	10.4	1.0	34.9	62.5	3.5	9.7	1.4	20.3	9.9	10.0	39.4	50.0	0.0	0.0	0.2	14.3	12.1	11.8	41.9	57.3
3109	287	14:0:35.41	139.4°S	62.0°E	8:31	21.9	36.1	59.0	1.3	9.8	0.0	48.7	1.7	8.7	0.3	19.5	3.1	9.0	30.9	59.6	9.2	8.7	31.1	60.2	1.8	9.0	0.4	19.8	6.5	10.3	39.9	50.0	0.0	0.0	0.2	14.3	12.1	11.8	41.9	57.3			
3268	312	14:0:35.42	139.4°S	62.0°E	17:18	14.2	9.8	22.7	28.9	60.3	2.3	9.3	0.0	0.0	50.2	2.0	10.4	0.1	20.1	15.0	0.7	30.6	61.3	15.7	11.7	30.9	62.0	2.1	10.9	0.7	20.7	10.2	10.0	39.8	53.8	0.0	0.0	0.2	14.3	12.1	11.8	41.9	57.3

⋮

【図 1 2】

ダスト	加速度	近傍100mの分散値		近傍100mの平均値	
		ダスト	加速度	ダスト	加速度
0.3	0.0	0.2	14.3	12.1	9.8

フロントページの続き

(72)発明者 川崎 仁史
東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内
(72)発明者 佐藤 浩史
東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内

審査官 圓道 浩史

(56)参考文献 特開2007-279887(JP,A)
特開2005-078277(JP,A)
米国特許出願公開第2008/0279434(US,A1)
特開2014-063432(JP,A)
倉沢 央 Hisashi KURASAWA, 不完全センサデータの重回帰における変数選択手法 Variable Selection Method in Multiple Regression with Incomplete Sensor Data, 電子情報通信学会技術研究報告 Vol.112 No.242 IEICE Technical Report, 日本, 一般社団法人電子情報通信学会 The Institute of Electronics, Information and Communication Engineers, 2012年10月10日, 第112巻、第242号, p.149-154
宮崎 陽司 Yoji Miyazaki, FIT2010 第9回情報科学技術フォーラム 講演論文集 第4分冊 査読付き論文・一般論文 ネットワーク・セキュリティ ユビキタス・モバイルコンピューティング 教育・人文科学 情報システム Forum on Information Technology 2010, 日本, 社団法人電子情報通信学会, 2010年 8月20日, p.431-434
内田 稜真 RYOMA UCHIDA, ウェアラブルセンサを用いた行動認識技術の欠損特徴量補完による耐故障性向上 Improving Fault Tolerance of Wearable Sensor-based Activity Recognition Technique by Completing Defective Feature Value, 情報処理学会研究報告 2012(平成24)年度 1 [CD-ROM], 日本, 一般社団法人情報処理学会, 2012年 6月 15日, p.1-8

(58)調査した分野(Int.Cl., DB名)

G 0 6 F 1 7 / 0 0 - 1 7 / 1 8
1 9 / 0 0
G 0 6 N 5 / 0 0 - 7 / 0 6
G 0 6 Q 1 0 / 0 0 - 1 0 / 1 0
3 0 / 0 0 - 3 0 / 0 8
5 0 / 0 0 - 5 0 / 2 0
5 0 / 2 6 - 9 9 / 0 0