

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2006-79479

(P2006-79479A)

(43) 公開日 平成18年3月23日(2006.3.23)

(51) Int. Cl.
G06F 21/20 (2006.01)

F I
G06F 15/00 330A

テーマコード(参考)
5B085

審査請求 有 請求項の数 5 O L (全 19 頁)

(21) 出願番号 特願2004-264758 (P2004-264758)
(22) 出願日 平成16年9月10日(2004.9.10)

特許法第30条第1項適用申請有り 2004年7月6日 社団法人電子情報通信学会発行の「電子情報通信学会技術研究報告 信学技報Vol. 104 No. 176」に発表

(71) 出願人 503360115
独立行政法人科学技術振興機構
埼玉県川口市本町4丁目1番8号
(74) 代理人 100091443
弁理士 西浦 ▲嗣▼晴
(72) 発明者 岡 瑞起
茨城県つくば市天久保3-7-2 シンセイマンション212
(72) 発明者 加藤 和彦
茨城県つくば市吾妻4-8-2
Fターム(参考) 5B085 AC11 AE00

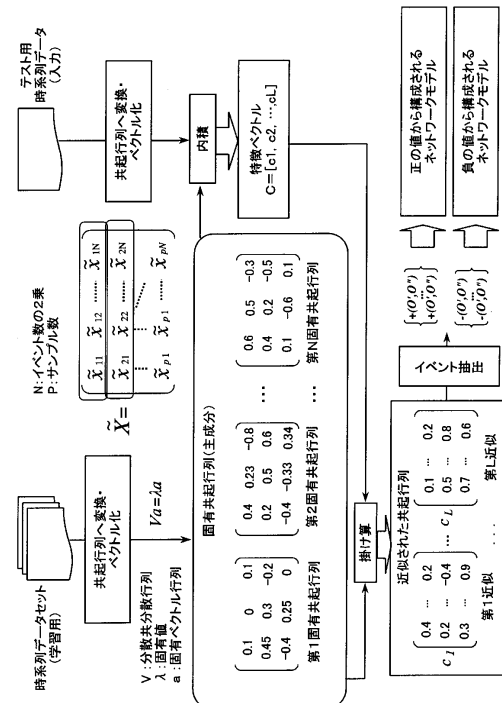
(54) 【発明の名称】 時系列データ判定方法

(57) 【要約】

【課題】従来よりも判定精度を高めることができる時系列データ判定方法を提供する。

【解決手段】 学習用時系列データから作成した多層ネットワークモデルと、テスト時系列データから作成した多層ネットワークモデルとの類似度を計算して、テスト時系列データが1以上のカテゴリに属するか否かを判定する。それぞれの層における多層ネットワークモデルは、特徴ベクトルの要素とこれに対応する固有共起行列とを掛け合わせて作成する。

【選択図】 図4



【特許請求の範囲】

【請求項 1】

複数種類のイベントを含んで構成される時系列データが所定の 1 以上のカテゴリに属するものであるか否かを判定する時系列データ判定方法であって、

予め学習用の複数の時系列データをそれぞれ予め定めたデータ長さのウィンドウで切り出して複数のウィンドウ・データを取り出すウィンドウ・データ取出ステップと、

前記ウィンドウ・データから前記データ長よりも短いデータ長を有する複数のスコープ・データを時間的なずれを持って順次抽出するスコープ・データ抽出ステップと、

前記複数のウィンドウ・データを複数の前記スコープ・データに基づいて前記ウィンドウ・データに含まれる前記複数種類のイベント相互間の時系列で見た関連性の強さを示す複数の共起行列に変換する共起行列変換ステップと、

前記複数の共起行列を入力として統計的特徴抽出方法により特徴ベクトルを求めるための基礎となる固有共起行列群を決定する固有共起行列群決定ステップと、

前記 1 以上のカテゴリを含む 1 以上のプロファイル学習用時系列データに対して前記ウィンドウ・データ取出ステップ、前記スコープ・データ抽出ステップ及び前記共起行列変換ステップと同様のステップをそれぞれ実施して、前記 1 以上のプロファイル学習用時系列データを 1 以上のプロファイル用共起行列に変換するプロファイル用共起行列変換ステップと、

前記 1 以上のプロファイル用共起行列と前記固有共起行列群とに基づいて前記 1 以上のプロファイル学習用時系列データについての 1 以上の判定用特徴ベクトルを抽出する判定用特徴ベクトル抽出ステップと、

テストの対象となるテスト時系列データに対して前記ウィンドウ・データ取出ステップ、前記スコープ・データ抽出ステップ及び前記共起行列変換ステップと同様のステップを実施して、前記テスト時系列データをテスト用共起行列に変換するテスト用共起行列変換ステップと、

前記テスト用共起行列と前記固有共起行列群とに基づいて前記テスト用時系列データについてのテスト用特徴ベクトルを抽出するテスト用特徴ベクトル抽出ステップと、

前記判定用特徴ベクトルと前記固有共起行列群を構成する複数の固有共起行列との掛け算に基づいて、前記複数の固有共起行列の次元を小さくした複数の判定用近似共起行列を取得する判定用近似共起行列取得ステップと、

前記複数の判定用近似共起行列からイベント抽出を行って判定用多層ネットワークモデルを生成する判定用多層ネットワークモデル生成ステップと、

前記テスト用特徴ベクトルと前記固有共起行列群を構成する複数の固有共起行列とを掛け算することにより、前記複数の固有共起行列の次元を小さくした複数のテスト用近似共起行列を取得するテスト用近似共起行列取得ステップと、

前記複数のテスト用近似共起行列からイベント抽出を行ってテスト用多層ネットワークモデルを生成するテスト用多層ネットワークモデル生成ステップと、

前記判定用多層ネットワークモデルと前記テスト用多層ネットワークモデルとに基づいて、前記テスト時系列データが前記 1 以上のカテゴリに属するか否かを判定する判定ステップとからなる時系列データ判定方法。

【請求項 2】

前記判定ステップでは、前記判定用多層ネットワークモデルと前記テスト用多層ネットワークモデルとの類似度に基づいて前記テスト時系列データが前記 1 以上のカテゴリに属するか否かを判定することを特徴とする請求項 1 に記載の時系列データ判定方法。

【請求項 3】

前記判定ステップでは、前記類似度が予め定めた閾値より大きいか否かにより、前記テスト時系列データが前記 1 以上のカテゴリに属するか否かを判定することを特徴とする請求項 2 に記載の時系列データ判定方法。

【請求項 4】

前記判定用多層ネットワークモデル及び前記テスト用多層ネットワークモデルは、それ

ぞれ共起性が正の値から構成されるネットワークモデルと、共起性が負の値から構成されるネットワークモデルとからなり、

前記判定ステップでは、前記共起性が正の値から構成されるネットワークモデルと、前記共起性が負の値から構成されるネットワークモデルの少なくとも一方を用いて、前記類似度を求めることを特徴とする請求項 2 または 3 に記載の時系列データ判定方法。

【請求項 5】

請求項 1 乃至 4 のいずれか 1 項に記載の時系列データ判定方法を用いて、コンピュータシステムに入力される時系列データの異常を判別することを特徴とする時系列データ異常判別方法。

【発明の詳細な説明】

10

【技術分野】

【0001】

本発明は、時系列データが所定の 1 以上のカテゴリに属するものであるか否かを判定する時系列データ判定方法に関するものである。

【背景技術】

【0002】

ユーザのパスワードを盗み出し、そのユーザになりすまして不正にコンピュータを使用する、いわゆる「なりすまし」を検出するためには、コンピュータに入力される時系列データに異常があるか否か（入力された時系列データが、なりすまし者によって作成された時系列データであるか否か）を異常検知システムで検知することが効果的である。公知の異常検知システムでは、最初にユーザの典型的な行動を示すプロファイル（ユーザが作成した時系列データに現れる特徴）を作成する。そしてテスト対象である入力データ（時系列データ）のプロファイルをそのユーザのプロファイルと比較することにより、正常なユーザが作成した時系列データであるか、なりすまし者が作成した異常な時系列データであるかを識別する。

20

【0003】

典型的な検査対象となる入力データは、使用された UNIX（登録商標）コマンド、アクセスされたファイル等の時系列データ等である。入力された時系列データが、正常か異常かを識別する過程は 2 つのステップに分けられる。まず第 1 ステップにおいて、時系列データの特徴抽出を行う。そして第 2 ステップにおいて、抽出された特徴が正常か異常かを識別する。

30

【0004】

第 1 のステップの特徴抽出を行う代表的な従来手法には、ヒストグラム（Histogram）とエヌグラム（N-gram）とがある。ヒストグラム（Histogram）では、時系列データに現れる項目（イベント）の出現頻度ベクトルが抽出の対象となる特徴ベクトルとなる。また、エヌグラム（N-gram）では、連続する N 個の項目を 1 つの特徴とする [非特許文献 1 乃至 3]。

【0005】

また第 2 ステップの抽出された特徴を正常か異常かを識別する手法としては、これまでに様々な手法が提案されている。それらの中で代表的な手法には、ルールベース [非特許文献 4]、オートマトン [非特許文献 5]、ベイジアンネットワーク [非特許文献 6]、Naive ベース [非特許文献 7]、ニューラルネットワーク [非特許文献 8]、マルコフモデル [非特許文献 9]、隠れマルコフモデル [非特許文献 10] とがある。

40

【0006】

これら問題に対処するために、発明者は、ユーザの挙動の動的情報をとらえて時系列データの特徴を抽出する方法 [(Eigen Co - occurrence Matrix (ECM 手法))] を提案している [非特許文献 11]。この ECM では、時系列情報を考慮しながら、イベント間の関連付けを行う。この関連付けは、二項間イベントに着目し全ての二項間イベントの関連性を Co - occurrence Matrix（共起行列）として表現することにより行う。共起行列は、全ての二項間の関連性の強さがその距離と

50

出現頻度により表現されることになる。

【非特許文献1】Ye, X. Li, Q. Chen, S. M. Emran, 及びM. Xu著の「Probabilistic Techniques for Intrusion Detection Based on Computer Audit Data」IEEE Transactions of Systems Man and Cybernetics, Vol. 31, pp. 266 - 274, 2001

【非特許文献2】S. A. Hofmeyr, S. Forrest 及びA. Somayaji著の「Intrusion Detection using Sequences of System Calls」Journal of Computer Security, vol. 6, pp. 151 - 180, 1998

【非特許文献3】W. Lee 及びS. J. Stolfo著の「A framework for constructing features and models for intrusion detection systems」, Information and System Security, vol. 3, pp. 227 - 261, 2000

【非特許文献4】N. Habra, B. L. Charlier, A. Mounji 及びI. Mathieu著の「ASAX: Software Architecture and Rule-Based Language for Universal Audit Trail Analysis」In Proc. of European Symposium on Research in Computer Security (ESORICS), pp. 435 - 450, 1992

【非特許文献5】R. Sekar, M. Bendre 及びP. Bollineni著の「A Fast Automaton Based Method for Detecting Anomalous Program Behaviors」In Proceedings of the 2001 IEEE Symposium on Security and Privacy, pp. 144 - 155, Oakland, May 2001.

【非特許文献6】W. DuMouchel著の「Computer Intrusion Detection Based on Bayes Factors for Comparing Command Transition Probabilities」Technical Report TR91, National Institute of Statistical Sciences (NISS), 1999.

【非特許文献7】R. A. Maxion 及びT. N. Townsend. 著の「Masquerade Detection Using Truncated Command Lines」In Proc. of the International Conference on Dependable Systems and Networks (DSN-02), pp. 219 - 228, 2002.

【非特許文献8】A. K. Ghosh, A. Schwartzbard, 及びM. Schatz著の「A study in using neural networks for anomaly and misuse detection」In Proc. of USENIX Security Symposium, pp. 141 - 151, 1999.

【非特許文献9】J. S. Tan, K. M. C. 及びR. A. Maxion. Markov Chains 著の「Classifiers and Intrusion Detection」In Proc. of 14th IEEE Computer Security Foundations Workshop, pp. 206 - 219, 2001

【非特許文献10】C. Warrender, S. Forrest 及びB. A. Pearlmutter著の「Detecting Intrusions using System Calls: Alternative Data Models」In I

10

20

30

40

50

EEE Symposium on Security and Privacy, pp. 133 - 145, 1999.

【非特許文献11】2004年3月1日に日本ソフトウェア学会主催の「第7回プログラミング及び応用のシステムに関するワークショップ」で配布された、岡瑞起、大山恵弘及び加藤和彦著の「Eigen Co-occurrence Matrix Method for Masquerade Detection」の論文

【発明の開示】

【発明が解決しようとする課題】

【0007】

しかしながらヒストグラム(Histogram)では、時系列データに現れる項目(イベント)の出現頻度ベクトルが特徴となる。また、エヌグラム(N-gram)では、連続するN個の項目を1つの特徴とする。しかしながらこれらの従来手法では、時系列データにおけるユーザの挙動の動的情報(時系列で見たユーザの挙動に関する情報即ちイベント時系列上に現れるイベントの種類とそれらの出現順で定まる各ユーザの特異的な特徴)が利用できない又は時系列データにおけるユーザの挙動の動的情報が失われるという問題や、単独もしくは隣接するイベントの特徴しか利用できない又は隣接するイベント間の特徴しか表現されないという問題がある。

10

【0008】

また発明者等が先に提案したECM手法を用いて正当なユーザとなりすまし者を識別するには、さらに共起行列をパターンとして扱い、統計的パターン認識手法を適用することが妥当と考えられる。最も簡単なパターン認識手法は、パターン間のマッチングに基づく手法である。しかし、共起行列そのものをパターンとして扱った場合、パターンの次元が膨大になってしまう。そのため、パターン間のマッチングでは、特徴を抽出し(情報圧縮にもなっている)、認識を行うことがより有効である。先に提案した具体的な方法では、共起行列から特徴ベクトルを求め、所定のベクトル識別関数を用いて時系列データと判定用特徴ベクトルとのユークリッド距離が閾値以内であるか否かによりその時系列データが1以上のカテゴリを含むか否か(正当なユーザが作成した時系列データであるか否か)を判定している。この先に提案した方法でも、所定の判定精度は確保することができる。しかしながら先に提案した判別方法では判定精度を高めることに限界があった。

20

【0009】

本発明の目的は、従来よりも判定精度を高めることができる時系列データ判定方法を提供することにある。

30

【0010】

本発明の別の目的は、時系列データに異常があるか否かを判別することができる時系列データ異常判別方法を提供することにある。

【課題を解決するための手段】

【0011】

本発明は、Eigen Co-occurrence Matrix(ECM)手法を開発したことを基礎としてなされたものである。このECM手法は、まず時系列情報を考慮しながら、時系列データに含まれるイベント間の関連付けを行う。この関連付けは、二つのイベント間の関連に着目し、全ての二項間イベントの関連性をCo-occurrence Matrix(共起行列)として表現することにより行う。共起行列は時系列データに現れる項目(イベント)間全ての関係性が表現することができる。これは、ヒストグラム(Histogram)やエヌグラム(N-gram)では表現することができなかった時系列データの特徴である。具体的な発明では、共起行列に対し主成分分析を行い、直交する主成分ベクトル空間を生成する。それぞれの共起行列は、主成分ベクトル空間上のベクトルとして特徴が抽出される。特徴をベクトルとして抽出することにより、様々なベクトル識別関数を利用することも可能になる。

40

【0012】

本発明の時系列データ判定方法は、複数種類のイベントを含んで構成される時系列データが所定の1以上のカテゴリに属するものであるか否かを特徴抽出方法と識別方法とを用

50

いて判定する。本発明では、特に、前記特徴抽出方法として、複数の時系列入力データを複数種類のイベントに含まれる二種類のイベント間の関連性を共起行列で表した行列データに変換したものをを用いる統計的特徴抽出方法を用いる。そして識別方法として統計的特徴抽出方法で抽出した特徴ベクトルを識別に利用するものを用いる。ここで複数種類のイベントとは、時系列データを構成する複数の項目を意味し、時系列データが複数のコマンドから構成されている場合には、その複数のコマンドがそれぞれイベントである。またカテゴリとは、上位概念で見れば時系列データの種別を意味する概念であって、下位の概念で見れば時系列データから得られる後述する特徴ベクトルの集合が属する種別である。例えば、ある時系列データがある正常であるか否かは、時系列データが予め定めた1以上のカテゴリに属するか否かにより判断することができる。なお特徴ベクトルとカテゴリとの関係で見れば、特徴ベクトルが存在する空間の部分領域に対応するものがカテゴリとなる。

10

【0013】

本発明の方法で採用する共起行列は、時系列データに現れる項目（イベント）間全ての関係性を表現することができる。言い替えると、共起行列は、全ての二項間の関連性の強さをその距離と出現頻度により表現する。したがって本発明によれば、時系列データに含まれる動的情報を利用して、時系列データが所定のカテゴリに属するか否かを従来よりも高い精度で判定することができる。

【0014】

複数の時系列入力データを共起行列で表した行列データに変換する際には、ウィンドウ・データ取出ステップと、スコープ・データ抽出ステップと、共起行列変換ステップとを実施する。ウィンドウ・データ取出ステップでは、時系列入力データをそれぞれ予め定めたデータ長さのウィンドウで切り出して複数のウィンドウ・データを取り出す。ウィンドウのデータ長さは、時系列データの長さに応じて定めればよい。スコープ・データ抽出ステップでは、ウィンドウ・データからウィンドウ・データのデータ長よりも短いデータ長を有する複数のスコープ・データをデータ列上において時間的なずれを持って順次抽出する。具体的なスコープ・データ抽出ステップでは、複数種類のイベントから選択した1つの種類のイベントがウィンドウ・データに含まれる位置を基準位置として1つの種類のイベントに対する1以上のスコープ・データを抽出することができる。また共起行列変換ステップでは、複数のウィンドウ・データを複数のスコープ・データに基づいてウィンドウ・データに含まれる複数種類のイベント相互間の時系列で見た関連性の強さを示す複数の共起行列に変換する。具体的な、共起行列変換ステップでは、1つの種類のイベントについての1以上のスコープ・データに含まれるその1つの種類のイベントまたは他の種類のイベントの数の合計値を、1つの種類のイベントに対する一つの種類のイベントの頻度とし、この頻度を1つの種類のイベントに対する一つの種類のイベントの関連性の強さを表示する値とする変換を行ってウィンドウ・データを共起行列に変換する。このようにして共起行列の変換を行うと、時系列で見たイベント相互間の関連性をより適格に示す共起行列を得ることができる。

20

30

【0015】

正当なユーザとなりすまし者を本発明の方法を利用して識別するには、さらに共起行列をパターンとして扱い、統計的パターン認識手法（識別方法）を適用することが妥当である。最も簡単なパターン認識手法（識別方法）は、パターン間のマッチングに基づく手法である。しかし共起行列そのものをパターンとして扱った場合、パターンの次元が膨大になってしまう。そのため、パターン間のマッチングでは、特徴を抽出し（情報圧縮にもなっている）、認識を行うことがより有効である。パターンから有効な特徴抽出を行うことにより、入力パターンの変動に対して頑健な認識結果が期待できる。そこで本発明のより具体的な方法では、特徴抽出方法として、主成分分析を用いて、共起行列からの特徴ベクトルの抽出に利用する。主成分分析はベクトル形式のデータを少数の特徴（主成分）で表すことを可能とする統計的特徴抽出方法である。なお主成分分析を用いた認識の成功例として、Turk等[M. Turk, A. Pentland, 「Eigenfaces f

40

50

or Recognition」Journal of Cognitive Neuroscience, vol 3, No. 1, 1991]が提案したEigenface (固有顔)による顔画像の認識が広く知られている。本発明の具体的方法では、共起行列 (Co-occurrence Matrix)を顔画像と見なしたところにユニークな着眼点がある。

【0016】

そこで複数種類のイベントを含んで構成される時系列データが所定の1以上のカテゴリに属するものであるか否かを判定する本発明の具体的な時系列データ判定方法では、前述のウィンドウ・データ取出ステップと、前述のスコープ・データ抽出ステップと、前述の共起行列変換ステップに加えて、更に固有共起行列群決定ステップと、プロファイル用共起行列変換ステップと、判定用特徴ベクトル抽出ステップと、テスト用共起行列変換ステップと、テスト用特徴ベクトル抽出ステップと、判定用近似共起行列取得ステップと、判定用多層ネットワークモデル生成ステップと、テスト用近似共起行列取得ステップと、テスト用多層ネットワークモデル生成ステップと、判定ステップとを用いる。

10

【0017】

固有共起行列群決定ステップでは、複数の共起行列を入力として主成分分析により特徴ベクトルを求めるための基礎となる固有共起行列群を決定する。またプロファイル用共起行列変換ステップでは、1以上のカテゴリを含む1以上のプロファイル学習用時系列データに対してウィンドウ・データ取出ステップ、スコープ・データ抽出ステップ及び共起行列変換ステップと同様のステップをそれぞれ実施して、1以上のプロファイル学習用時系列データを1以上のプロファイル用共起行列に変換する。また判定用特徴ベクトル抽出ステップでは、1以上のプロファイル用共起行列と固有共起行列群とに基づいて1以上のプロファイル学習用時系列データについての1以上の判定用特徴ベクトルを抽出する。更にテスト用共起行列変換ステップでは、テストの対象となるテスト時系列データに対してウィンドウ・データ取出ステップ、スコープ・データ抽出ステップ及び共起行列変換ステップと同様のステップを実施して、テスト時系列データをテスト用共起行列に変換する。またテスト用特徴ベクトル抽出ステップは、テスト用共起行列と固有共起行列群とに基づいてテスト用時系列データについてのテスト用特徴ベクトルを抽出する。

20

【0018】

更に判定用近似共起行列取得ステップでは、判定用特徴ベクトルと固有共起行列群を構成する複数の固有共起行列との掛け算に基づいて、複数の固有共起行列の次元を小さくした複数の判定用近似共起行列を取得する。また判定用多層ネットワークモデル生成ステップでは、複数の判定用近似共起行列からイベント抽出を行って判定用多層ネットワークモデルを生成する。またテスト用近似共起行列取得ステップでは、テスト用特徴ベクトルと固有共起行列群を構成する複数の固有共起行列とを掛け算することにより、複数の固有共起行列の次元を小さくした複数のテスト用近似共起行列を取得する。またテスト用多層ネットワークモデル生成ステップでは、複数のテスト用近似共起行列からイベント抽出を行ってテスト用多層ネットワークモデルを生成する。

30

【0019】

上記の各ステップを実行した後に判定ステップでは、判定用多層ネットワークモデル (多層ネットワークモデル)とテスト用多層ネットワークモデル (多層ネットワークモデル)とに基づいて、テスト時系列データが1以上のカテゴリに属するか否かを判定する。本発明の判別法を用いると、先に提案している判別法と比べて高い判定精度で判定することができる。

40

【0020】

具体的な判定ステップでは、例えば、判定用多層ネットワークモデルとテスト用多層ネットワークモデルとの類似度に基づいてテスト時系列データが前記1以上のカテゴリに属するか否かを判定することができる。その場合、判定ステップでは、類似度が予め定められた閾値より大きいかが否かにより、テスト時系列データが前記1以上のカテゴリに属するか否かを判定するのが好ましい。なお判定用多層ネットワークモデル及びテスト用多層ネット

50

ワークモデルは、それぞれ共起性が正の値から構成されるネットワークモデルと、共起性が負の値から構成されるネットワークモデルとから構成できる。この場合には、判定ステップでは、共起性が正の値から構成されるネットワークモデルと、共起性が負の値から構成されるネットワークモデルの少なくとも一方を用いて、類似度を求めればよい。なお正のネットワークは固有共起行列群を作成するのに用いた共起行列群の平均共起行列よりも強い特徴を表し、負のネットワークは、その平均共起行列よりも弱い特徴を表す。両方を用いると、平均共起行列よりも強い特徴か弱い特徴からの2つの視点から判定することができるので判定精度を更に高めることができる。

【0021】

また本発明の時系列データ判定方法を用いて、コンピュータシステムに入力される時系列データの異常を判別すると、従来よりも高い精度で異常な時系列データを判別することができる。

10

【発明の効果】

【0022】

本発明によれば、時系列データが所定のカテゴリを含むものであるか否かを従来よりも高い精度で判定することができる。

【発明を実施するための最良の形態】

【0023】

以下図面を参照して本発明の実施の形態を詳細に説明する。図1は、複数種類のイベントを含んで構成される時系列データが所定の1以上のカテゴリに属するものであるか否かを主成分分析法を用いて判定する本発明の時系列判定方法の実施の形態の一例を実施するためのプログラムにおいて、特徴ベクトルを求めるまでの構成を示す図である。なお特徴ベクトルを求めるまでの方法の基本は、既に発明者等が非特許文献11で公表している。図1においては、特徴ベクトルを得るために用いる固有共起行列群を得るための学習用の複数の時系列データと、プロファイル学習用の時系列データ(以下プロファイル学習用時系列データと言う)と、テストの対象となるテスト時系列データ(以下テスト用時系列データと言う)を共起行列に変換する。ここで共起行列とは、時系列データを構成する複数種類のイベントに含まれる二種類のイベント間の関連性を行列データに変換したものである。

20

【0024】

時系列データを共起行列に変換するステップについて説明する。図2は、複数の学習用時系列データ[この場合にはユーザ(コンピュータにアクセスして時系列データを送信してくる人または他のコンピュータ)1乃至ユーザ3からそれぞれ送られた3つの時系列データの]の構成の一例を示している。この例では、各ユーザからの時系列データは、20のコマンド(イベント)によってそれぞれ構成されている。後に説明するように、この実施の形態では、20のコマンドからなる時系列データを10のコマンド(データ長)を有するウィンドウで区切る(ウィンドウ・データ取出ステップ)。このウィンドウ・データ取出ステップでは、各時系列入力データをそれぞれ予め定めたデータ長(10個のコマンド分のデータ長)のウィンドウで切り出して2つのウィンドウ・データを取り出す。なおウィンドウのデータ長さは、時系列データの長さに応じて定めればよい。

30

40

【0025】

次に、ある区間の時系列データに現れる2つのイベント間の因果関係を表すために共起行列に変換する。共起行列のそれぞれの要素は、2つのイベント間の因果関係の強さを表すものである。共起行列を作成するために、ウィンドウサイズ w 、スコープサイズ s 、そしてイベントセット $B = \{b_1, b_2, b_3, \dots, b_m\}$ を定義する。ここで m は、イベント数を示す。ウィンドウサイズ w は、1つの特徴ベクトルを抽出するイベント時系列のサイズを決定し、スコープサイズ s は、2つのイベントの因果関係を考慮する間隔幅を決定する。図2に示すデータ例では、 w を10、 s を6と定義した。また、 B は、3人全ての学習用の時系列データ(ドメインデータ)に現れるユニークな8つのコマンド(イベント)($m = 8$)とする。8つのコマンドは、`cd`, `ls`, `less`, `emacs`, `gcc`, `gdb`, `mkdir`,

50

cpである。2つのイベント間の因果関係または関連性の強さは、イベント間の距離と、それらが現れる頻度により定義される。つまり、注目するイベントが、ウィンドウサイズ(10)の中で、スコープサイズ(6)以内に現れる頻度を数えることにより、イベント間の因果関係の強さを定義する。図2の例では、それぞれにユーザー人について2つの共起行列が作られることになる。図3のウィンドウ1におけるイベントcdとイベントlsの要素または頻度数7は、ウィンドウサイズ(10)で、スコープサイズ(6)以内に、lsがcdの後に7回現れたことを示している。イベントペア(cd ls)と{ls cd}が図3のウィンドウ1において最も大きな要素または頻度数を持つ。これはこの時系列において、これらのイベントは強い関係性があることを示している。共起行列は、時系列データに現れる全ての2つのイベント相互間の因果関係または関連性の強さを表現することになる。

10

【0026】

図3について、本発明との関係で、詳しく説明する。まず各ユーザの時系列データ毎に、図3に示すように、前述のウィンドウ・データから複数のスコープ・データを抽出する(スコープ・データ抽出ステップ)。このステップでは、ウィンドウ・データからウィンドウ・データのデータ長よりも短いデータ長を有する複数のスコープ・データをデータ上における時間的なずれを持って順次抽出する。この例では、6個のコマンド分のデータ長を有するスコープ・データを順次抽出している。具体的には、ウィンドウ・データを構成する10個のコマンドに含まれる複数種類のイベント(図3の場合には、cd, ls, le ss)から選択した1つの種類のイベント(例えばcd)が、ウィンドウ・データに含まれる位置を基準位置として1つの種類のイベントに対する1以上のスコープ・データを抽出する。図3の例で見れば、イベントcdに着目した場合、ウィンドウ1の先頭にあるイベントcdを含まずにこのcd(基準位置)より後の6個のコマンド(イベント)を第1のスコープ・データとして抽出し、次に先頭から6番目にあるイベントcdを含まずにこのcd(基準位置)より後の6個のコマンド(イベント)を第2のスコープ・データとして抽出する。なお図3の例のように、ウィンドウ1内に10個しかイベントが無い場合、第2のスコープ・データでは4個のイベントを抽出する。同様に、先頭から第8番目及び第9番目のイベントcdを基準位置にして第3及び第4のスコープ・データを抽出する。

20

【0027】

次に、ウィンドウ・データから抽出した複数のスコープ・データに基づいてそのウィンドウ・データに含まれる複数種類のイベント相互間の時系列で見た関連性の強さ(二つのイベントの相互間の関連強さ)を、関連性を見る二つのイベントが現れる頻度と距離として表現する。例えば、1つの種類のイベントcdについての1以上(図3の場合には4つ)のスコープ・データに含まれる1つの種類のイベント(図3の場合には同じ種類のcd)の数の合計値を、1つの種類のイベントに対する一つの種類のイベントの頻度とする。そして、この頻度を1つの種類のイベントに対する一つの種類のイベントの関連性の強さを表示する値とする変換を行ってウィンドウ・データを共起行列に変換する。図3の例において、ウィンドウ1中のイベントcdとイベントcdとの間の関連性を頻度として見る。前述の第1のスコープ・データ中には、1つのcdが含まれており、第2のスコープ・データ中には2つのcdが含まれており、第3のスコープ・データ中には1つのcdが含まれており、第4のスコープ・データ中にはcdは含まれない。したがってイベントcdに対するイベントcdの頻度は、 $1 + 2 + 1 + 0 = 4$ と計算できる。同様にしてイベントcdに対するイベントlsの関連性についてみれば、前述の第1のスコープ・データ中には、3つのlsが含まれており、第2のスコープ・データ中には2つのlsが含まれており、第3のスコープ・データ中には1つのlsが含まれており、第4のスコープ・データ中には1つのlsが含まれている。したがってイベントcdに対するイベントlsの頻度は、 $3 + 2 + 1 + 1 = 7$ と計算できる。これらの頻度には、スコープ・データを設定することにより、時間または距離の関係即ち時系列データに含まれる動的情報が含まれることになる。図3の右側領域には、ウィンドウ1及び2をそれぞれ共起行列に変換した行列データが示されている。このように時系列データを共起行列で表現すると、人間の流動的な

30

40

50

行動のモデル化が可能になる。

【0028】

正当なユーザとなりすまし者を本発明の方法を利用して識別するには、共起行列をパターンとして扱い、統計的特徴抽出方法として主成分分析を用いて特徴ベクトルを求め、その後特徴ベクトルを識別に利用して識別を実行する。主成分分析はベクトル形式のデータを少数の特徴（主成分）で表すことを可能とする統計的特徴抽出方法であり、主成分分析とは多変量で表されるデータの統計から、一次結合で表現される新たな変量を構成し、互いに無相関な「主成分」に要約する手法である。本実施の形態では、共起行列を先に述べた Turk 等が提案した Eigenface（固有顔）による顔画像と見なしている。そこで本出願においては、本発明の時系列データ判定方法を Eigen Co-occurrence Matrix (ECM) 手法と呼ぶ。 10

【0029】

図1に示すように、時系列データから、固有共起行列群を作成する学習用の時系列データを選びこれをドメインデータとする。1つのウィンドウから変換した共起行列を前述の M. Turk 等が発表した Eigenface（固有顔）における顔画像と見なし、Eigenface に対応する Eigen Co-occurrence Matrix（固有共起行列）を作成する。主成分分析により、固有値とそれに対応する固有ベクトルが得られる。そして固有値を降順に並べ、それと対応する固有ベクトルを上から N 個選択し、行列化し固有共起行列群とする。

【0030】

共起行列からの主成分分析を用いた特徴ベクトル抽出は次に述べる手順で行う。まず学習用の時系列データから得た p 枚の学習用の共起行列のうち i 番目の共起行列を、各要素の値を並べた N 次元のベクトル x_i として表現する。ここで p はサンプル数であり、N はイベント数の 2 乗である。p 枚の共起行列の平均ベクトルを平均共起行列として下記の式で求める。ここで平均共起行列は、イベントペア（2 項間）の関係性を示す。 20

【数1】

$$\bar{x} = \frac{1}{p} \sum_{i=1}^p x_i$$

30

【0031】

そして各共起行列から平均共起行列（平均ベクトル）を引いたベクトルを

【数2】

$$\tilde{x}_j = x_j - \bar{x}$$

【0032】

で表す。この平均共起行列を引く意味は、座標軸を原点に設定するためである。そして各共起行列から平均共起行列（ $m \times m$ 行列）を引き、ベクトル化した（ $m \times m$ の行列を m^2 次元の縦ベクトルにする）共起行列の集合を行列

【数3】

$$\tilde{X} = [\tilde{x}_1, \dots, \tilde{x}_p]$$

40

【0033】

で表す。この行列とその転置行列をかけた行列が図1における共分散行列（ $m^2 \times m^2$ 行列）である。

【0034】

次に、学習用の共起行列の集合を最適に近似する正規直交基底 a を、[数3] で表した行列 X の共分散行列の固有ベクトルで構成する。そのために共分散行列から固有値及び固有ベクトルを計算する（ $m^2 \times m^2$ 行列の固有ベクトルを計算）する。ここで固有値は、 50

特徴の強さを表す。また固有ベクトルは、お互いに無相関な特徴の軸を表している。このとき、 a の各固有ベクトル a_l を、固有共起行列 (Eigen co - occur ren ce matrix) とし、その集合を固有共起行列群 (主成分) とする。

【0035】

具体的には、固有値を降順にソートし、それらに対応する固有ベクトルを得る (m^2 個の固有ベクトルのうち L 個のみ選択する。固有値によって、固有ベクトルをソートすることにより、特徴の強い軸を上から順番に取り出すことができる。 L 個の固有ベクトルをそれぞれ行列化し (m^2 次元のベクトルを $m \times m$ の行列にする)、これを固有共起行列群とする。ここで固有共起行列 x に対する特徴ベクトル (A) (または主成分スコア C) を縦ベクトル化した共起行列 x と正規直交基底 a の内積を計算することにより求める。特徴ベクトルの各成分 c_1, c_2, \dots, c_L は、共起行列 x を表現するための各固有共起行列の貢献度を表すことになる。本実施の形態のように、特徴ベクトルを共起行列から抽出した場合、様々なベクトル空間手法を用いた特徴ベクトルの識別に使用することができる。

10

【0036】

本発明の時系列データの判定方法と関係する部分について以下に説明する。判定方法では、前述の共起行列の変換で用いたウィンドウ・データ取出ステップと、前述のスコープ・データ抽出ステップと、前述の共起行列変換ステップに加えて、更に固有共起行列決定ステップと、プロファイル用共起行列変換ステップと、判定用特徴ベクトル抽出ステップと、テスト用共起行列変換ステップと、テスト用特徴ベクトル抽出ステップと、判定用近似共起行列取得ステップと、判定用多層ネットワークモデル生成ステップと、テスト用近似共起行列取得ステップと、テスト用多層ネットワークモデル生成ステップと、判定ステップとを用いる。

20

【0037】

まず固有共起行列決定ステップでは、前述のようにして複数の共起行列 (学習用の時系列データを共起行列に変換したもの) を入力として主成分分析により特徴ベクトルを求めるための基礎となる固有共起行列群 (固有共起行列の集合即ち主成分) を決定する。

【0038】

そしてプロファイル用共起行列変換ステップでは、1以上のカテゴリを含む1以上のプロファイル学習用時系列データに対して先に説明したのと同様のウィンドウ・データ取出ステップ、スコープ・データ抽出ステップ及び共起行列変換ステップと同様のステップをそれぞれ実施して、1以上のプロファイル学習用時系列データを1以上のプロファイル用共起行列に変換する。ここでプロファイル学習用時系列データとしては、正常なユーザが作成したものであることが明確に判っている時系列データを用いる。学習用の時系列データからこのプロファイル学習用時系列データを選んでもよいのは勿論である。あるコンピュータにアクセスするユーザが100人いれば、その100人が作成した時系列データをプロファイル学習用時系列データとしてそれぞれプロファイル用共起行列に変換する。

30

【0039】

次に判定用特徴ベクトル抽出ステップでは、プロファイル用共起行列と固有共起行列群とに基づいて各プロファイル学習用時系列データについての判定用特徴ベクトルを抽出する。このようにして抽出した判定用特徴ベクトルは、事前にコンピュータのメモリに記憶しておく。なお図1には、特にプロファイル学習用時系列データについては記載していないが、テスト用時系列データと同じルートで共起行列に変換し、その特徴ベクトルを求める。

40

【0040】

次に、テスト用共起行列変換ステップでは、テストの対象となるテスト時系列データに対してウィンドウ・データ取出ステップ、スコープ・データ抽出ステップ及び共起行列変換ステップと同様のステップを実施して、テスト時系列データをテスト用共起行列に変換する。また、テスト用特徴ベクトル抽出ステップは、テスト用共起行列と固有共起行列とに基づいてテスト用時系列データについてのテスト用特徴ベクトルを抽出する。なお、テスト用特徴ベクトルを抽出する際には、図1に示すようにテスト用共起行列から平均共起

50

行列を引いたものをベクトル化したものと先に求めた固有共起行列群をベクトル化したものとの内積を求める。

【 0 0 4 1 】

図 4 は、本実施の形態の判定ステップを実行するために用いるプログラムの構成部分を示したブロック図である。各特徴ベクトルを算出するまでは、図 1 に説明した通りである。判定ステップを実行する前に、先に求めて記憶してある判定用特徴ベクトルとテスト用特徴ベクトルとに基づいて、判定用近似共起行列取得ステップと、判定用多層ネットワークモデル生成ステップと、テスト用近似共起行列取得ステップと、テスト用多層ネットワークモデル生成ステップを実行する。

【 0 0 4 2 】

判定用近似共起行列取得ステップでは、判定用特徴ベクトルと固有共起行列群を構成する複数の固有共起行列（第 1 ～ 第 L の固有共起行列）との掛け算に基づいて、複数の判定用近似共起行列（第 1 ～ 第 L の近似共起行列：但し L は正の整数）を取得する。そして判定用多層ネットワークモデル生成ステップでは、複数の判定用近似共起行列（第 1 ～ 第 L の近似共起行列）からイベント抽出を行って判定用多層ネットワークモデルを生成する。ここでイベント抽出とは、閾値 h を定め、正の値から構成するネットワークにおいては、閾値 h よりも大きい値のみを用いてネットワークを構成すること、また負の値から構成するネットワークにおいては、閾値 - h よりも小さい値のみを用いてネットワークを構成することである。図 4 において、「イベント抽出」の後に示された 2 つの行列が、それぞれ多層ネットワークの正の値から構成されるネットワークモデルと、負の値から構成される

10

20

【 0 0 4 3 】

またテスト用近似共起行列取得ステップでも、前述と同様にして、先に求めたテスト用特徴ベクトルと固有共起行列群を構成する複数の固有共起行列（第 1 ～ 第 L の固有共起行列）とを掛け算することにより、複数のテスト用近似共起行列を取得する。そしてテスト用多層ネットワークモデル生成ステップでは、複数のテスト用近似共起行列からイベント抽出を行ってテスト用多層ネットワークモデルを生成する。

【 0 0 4 4 】

多層ネットワークモデルを生成するまでのステップをより具体的なレベルで説明する。まず固有共起行列の次元（個数）L（L = 1, …, N）を小さくすることにより、固有共起行列 a と特徴ベクトル c を用いて、もとの共起行列を下記 [数 4] のように低次元（少ない個数）で表現する。

30

【 数 4 】

$$\hat{x} = \sum_{i=1}^L c_i a_i \quad \text{for } (L=1, \dots, N)$$

【 0 0 4 5 】

また、i 番目の共起行列を下記 [数 5] を用いて生成し、第 i 層のネットワークを抽出することにより多層ネットワーク表現が可能である。下記 [数 5] の行列 z_i を近似共起行列を言う。固有共起行列群を a₁, a₂, …, a_L とし、特徴ベクトルを c₁, c₂, …, c_L とした場合、1 近似ネットワークは、c₁ × a₁ であり、2 近似ネットワークは c₁ × a₁ + c₂ × a₂ であり、L 近似ネットワークは $\sum c_i \times a_i$ となる。また多層ネットワークのそれぞれの層のネットワークは、固有共起行列群 a₁, a₂, …, a_L と特徴ベクトル c₁, c₂, …, c_L とした場合において、それぞれ対応する固有共起行列 a_i と特徴ベクトルの要素 c_i を掛けて生成する行列 a_i × c_i から得られるネットワークを意味する。

40

【数5】

$$z_i = c_i a_i \quad \text{for } (i=1, \dots, N)$$

【0046】

それぞれの層のネットワークは、もとの共起行列のサブネットワークではなく、固有共起行列から発生する全体の構造をもつネットワークである。

【0047】

さらに行列 z_i を下記【数6】に示すように、正 ($x(i)$) と負 ($y(i)$) の要素に分離して、それぞれから1つ1つのネットワークを構成できる。

【数6】

$$z_i = c_i a_i = x(i) + y(i) \quad \text{for } (L=1, \dots, N)$$

【0048】

正の要素から成る行列 ($x(i)$) の作るネットワークは共起性が正の値で、(入力 - 平均) の行列を再構成するのに寄与する。また負の要素から成る行列 ($y(i)$) は同じく負の値で、再構成に寄与するという違いがある。上記の多層ネットワークモデル生成手法を判定用多層ネットワークモデル及びテスト用多層ネットワークモデルの生成の両方で実施する。共起行列を隣接行列として捉えると、共起行列からのネットワーク構造が抽出される。図5に一つの共起行列をネットワーク構造として表現した結果を示す。このようなネットワーク構造にして、後の判定ステップを実施すると検知率は向上する。

【0049】

上記の各ステップを実行した後に判定ステップでは、判定用多層ネットワークモデルとテスト用多層ネットワークモデルとに基づいて、テスト時系列データが1以上のカテゴリに属するか否かを判定する。具体的な判定ステップでは、例えば、判定用多層ネットワークモデルとテスト用多層ネットワークモデルとの類似度に基づいてテスト時系列データが前記1以上のカテゴリに属するか否かを判定する。その場合、判定ステップでは、類似度が予め定めた閾値より大きいか否かにより、テスト時系列データが前記1以上のカテゴリに属するか否かを判定する。

【0050】

判定用多層ネットワークモデルとテスト用多層ネットワークモデルとの間の類似度は、各層のネットワークモデルが共有するサブネットワーク数 (E 個のエッジから構成されるネットワークを1つのサブネットワークとする) を計算し、それらを足した数とする。

【0051】

図6は、エッジが1個の場合とエッジが2個の場合のサブネットワークの例を示している。そして図7に示すように、判定用多層ネットワークモデルには、第1層から第L層までのネットワークがある。そして入力用多層ネットワークモデルにも第1層から第L層までのネットワークがある。これら各層の間でネットワーク間の類似度の計算をする。図8は、一つの層の判定用ネットワークモデルとテスト用ネットワークモデルとの間の類似度の計算例を示している。この例では、 $E=1$ とし、2つのネットワーク間のサブネットワーク数を数えて類似度を計算する。6つのサブネットワークを共有していることから、この場合には類似度は6となる。

【0052】

次に実際に上記実施の形態の方法を実行するプログラムを異常検知システムに実装し、実際のUNIX (登録商標) コマンド時系列のログデータにおいて正常なユーザとなりすまし者を識別する実験を行った結果について説明する。実験には Schonlau 等 (M. Schonlau, W. Dumonchel, W. H. Ju, A. F. Karr, M. Theus, Y. Vardi, Computer intrusion: Detecting masquerades, Statistica Science, vol.

10

20

30

40

50

16, no. 1, pp. 58 - 74, 2001.) が提供しているUNIX (登録商標) コマンドのデータを用いた。Schonlau等のデータには、1人のユーザにつき、15,000のUNIX (登録商標) コマンドの履歴が提供されており、それが50人分用意されている。彼らのデータには、プライバシーの理由から引数、フラグやエイリアスの情報は含まれていない。15,000コマンドのうち最初の5000コマンドは、正規のユーザのコマンドで構成されており、残りの10,000コマンドになりすましのデータが挿入されている。最初の5000コマンドを学習データ、残りの10,000コマンドをテストデータとした。

【0053】

検査対象は100コマンドずつとし、5,000コマンドの学習データを100コマンドごとのウィンドウに分け、それぞれから上記ECM手法を用いて特徴抽出を行う。50人全てのユーザの学習データを用い(50×50=2500ウィンドウ)、固有共起行列を作成した。そのうち降順に並べられた最初の50個の固有値に対応する50個の固有ベクトルを固有共起行列として用いた(L=50)。

【0054】

学習データとして、各ユーザの学習データから100コマンドを1ウィンドウの単位とし、上記実施の形態の方法を用いてネットワークモデルに変換した。

【0055】

テストデータとして、各ユーザのテストデータから同様に100コマンドを1ウィンドウ単位とし、上記実施の形態を用いてネットワークモデルに変換した。

【0056】

学習データネットワークモデルセットをSとし、それとテストデータseq_iのネットワークモデルの類似度を計算し、「正常」か「異常」であるかを判断する。seq_iの学習データとの類似度は、[数7]で表わされるように、学習データそれぞれのネットワークモデルとの類似度で1番大きい値とする。

【数7】

$$Sim(seq_i, S) = \max_{seq_j \in S} \{Sim(seq_i, seq_j)\}$$

【0057】

ここで、ネットワークモデルの類似度Sim(seq_i, seq_j)は、

【数8】

$$Sim(seq_i, seq_j) = \sum_k^N \delta(T_k(i), T_k(j))$$

【0058】

のように定義した。T_k(x)はseq_xがk層においてつくるネットワークモデルを示し、(T_k(i), T_k(j))は、ネットワークモデルT_k(i)とT_k(j)が同一層同士における部分ネットワークの一致数の和を示している。各層におけるネットワークは、対応する近似共起行列から値の大きい順に30個ノードを取り出し、構成した。また、エッジが2つ繋がっているネットワークを1つのサブネットワークと捉えた。

【0059】

ユーザiごとに、テストデータseq_iが「正常」であるか「異常」であるか判断する類似度のしきい値i₀を設け、Sim(seq_i, S)がしきい値i₀よりも大きければ正常、小さければ異常と判断する。しきい値i₀を変化させることにより検知率(異常な実行を異常と判断する)と誤検知率(正常な実行を異常と判断する)が変化する。

【0060】

実験の評価にはReceiver Operating Characteristic (ROC)カーブを用いた。ROCカーブとは、縦軸に検知率、横軸に誤検知率をとり

、しきい値を変化させたときの結果をプロットしたシステムの精度を表すグラフである。プロット点が図の左上に近ければ近いほど、誤検知率が低く、検知率が高いことを示し、性能が良いことを表す。

【0061】

Schonlau \bar{r} (M. Schonlau, W. Dumonchel, W. H. Ju, A. F. Karr, M. Theus, Y. Vardi, Computer Intrusion: Detecting masquerades, Statistica Science, vol. 16, no. 1, pp. 58-74, 2001.) と Maxion \bar{r} (R. A. Maxion, T. N. Townsend, Masquerade Detection Using Truncated Command Lines, Proc. International Conference on Dependable Systems and Networks (DSN-02), pp. 219-228, Washington, 2002.) は、本実験で使用した同様のデータセットに対し Bayes 1-Step Markov, Hybrid Multi-Step Markov, IPAM, Uniqueness, Sequence-Match, Compression, and Naive Bayes と呼ばれる手法を適用している。本実施の形態の方法を用いてユーザ i ごとに、しきい値 i を変化させ、50人分の結果をまとめた結果を彼らの結果とともに、図6に示す。図6の結果から分かるように、本実施の形態の方法が最も高い検知率の中で、最も低い誤検知率を示しており、本発明の有効性が確認できた。

10

20

【0062】

また図7は、先に発明者等が提案した特徴ベクトルを用いた場合の検知率及び誤検知率と本実施の形態の方法を用いた場合の検知率及び誤検知率を示している。図7においてAは、先に提案した特徴ベクトルを判定に利用した場合の検知率及び誤検知率を示しており、Bは多層ネットワークを用いる本実施の形態の方法を判定に利用した場合の検知率及び誤検知率を示している。図7から分かるように、本実施の形態の方法が、先に提案した方法よりも、高い検知率と低い誤検知率を示していることが分かる。

【0063】

また本発明の時系列データ判定方法を用いて、コンピュータシステムに入力される時系列データの異常を判別すると、従来よりも高い精度で異常な時系列データを判別することが

30

【図面の簡単な説明】

【0064】

【図1】複数種類のイベントを含んで構成される時系列データが所定の1以上のカテゴリに属するものであるか否かを主成分分析法を用いて判定する本発明の時系列判定方法の実施の形態の一例を実施するためのプログラムの構成を示す図である。

【図2】複数のユーザ1乃至ユーザ3からそれぞれ送られた3つの時系列データの構成の一例を示している。

【図3】共起行列の変換を説明するために用いる図である。

【図4】本実施の形態の判定ステップを実行するために用いるプログラムの構成部分を示したブロック図である。

40

【図5】一つの共起行列をネットワーク構造として表現した結果を示す。

【図6】エッジが1個の場合とエッジが2個の場合のサブネットワークの例を示す図である。

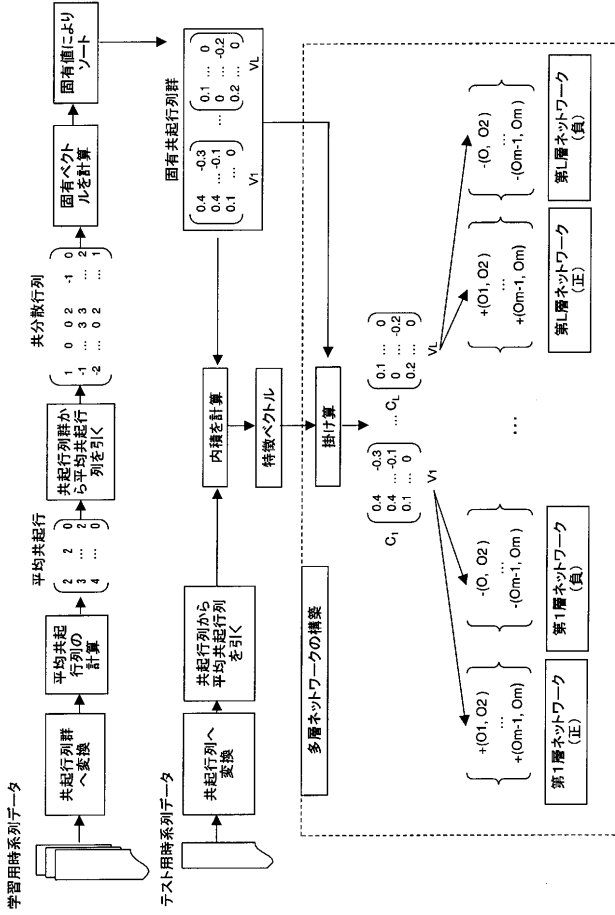
【図7】層間で類似度の計算をすることを示す図である。

【図8】一つの層の判定用ネットワークモデルとテスト用ネットワークモデルとの間の類似度の計算例を示す図である。

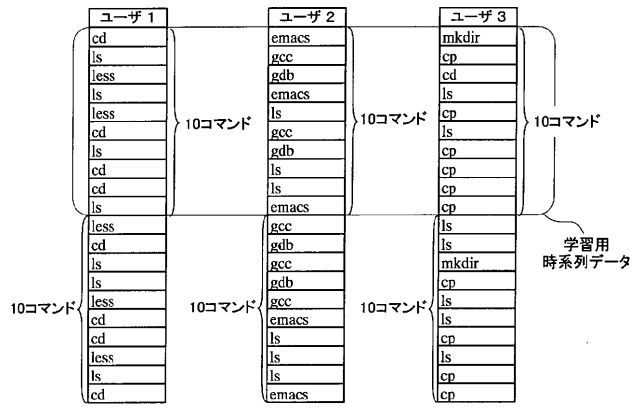
【図9】実験における検知率と誤検知率の関係を示す図である。

【図10】判定方法の相違による検知率と誤検知率の関係を示す図である。

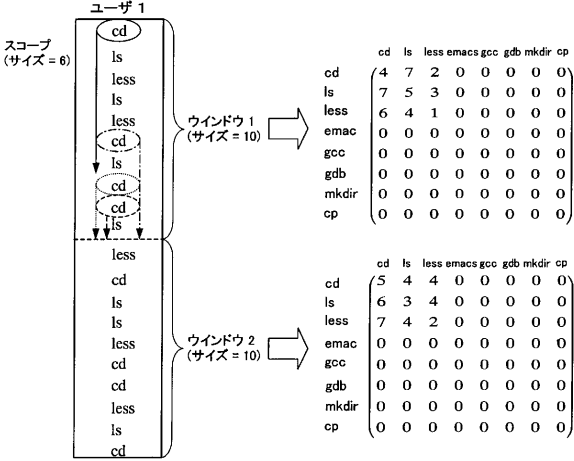
【図1】



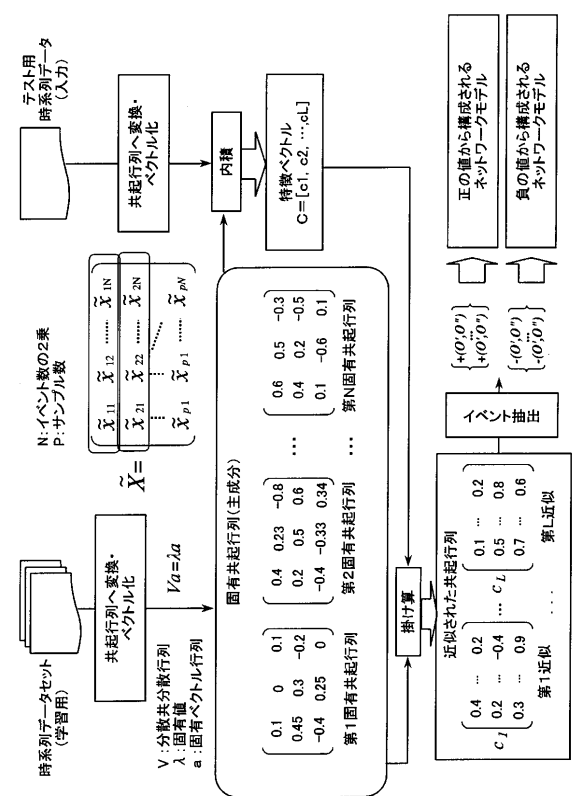
【図2】



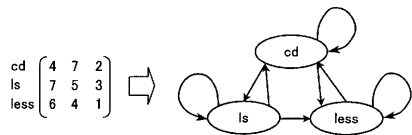
【図3】



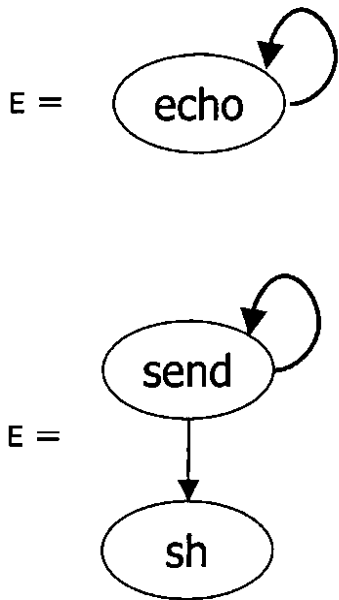
【図4】



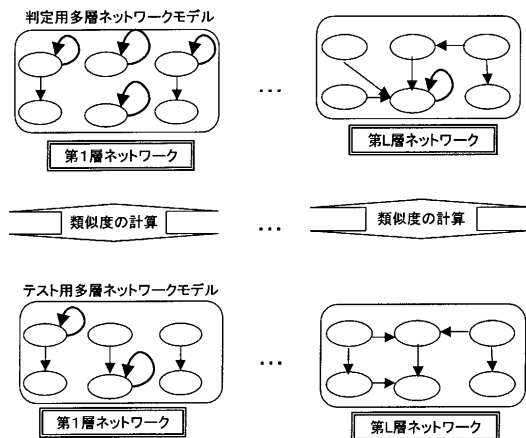
【 図 5 】



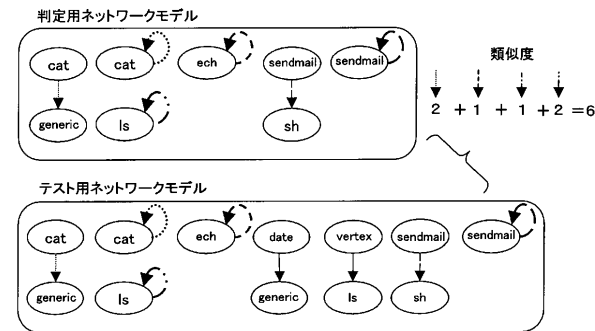
【 図 6 】



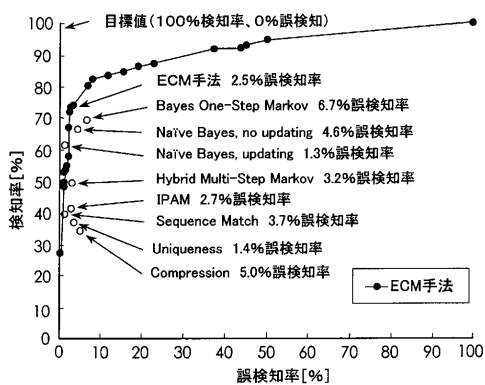
【 図 7 】



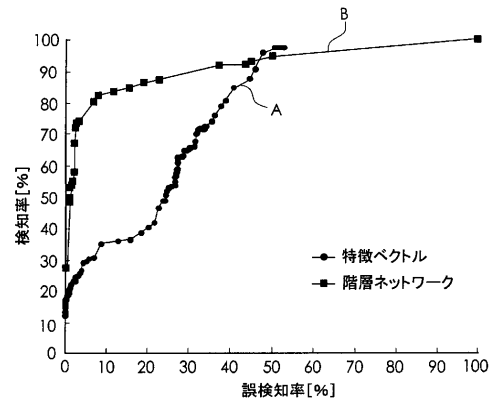
【 図 8 】



【 図 9 】



【 図 10 】



【手続補正書】

【提出日】平成16年10月7日(2004.10.7)

【手続補正1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

複数種類のイベントを含んで構成される時系列データが所定の1以上のカテゴリに属するものであるか否かを判定する時系列データ判定方法であって、

予め学習用の複数の時系列データをそれぞれ予め定めたデータ長のウィンドウで切り出して複数のウィンドウ・データを取り出すウィンドウ・データ取出ステップと、

前記ウィンドウ・データから前記データ長よりも短いデータ長を有する複数のスコープ・データを時間的なずれを持って順次抽出するスコープ・データ抽出ステップと、

前記複数のウィンドウ・データを複数の前記スコープ・データに基づいて前記ウィンドウ・データに含まれる前記複数種類のイベント相互間の時系列で見た関連性の強さを示す複数の共起行列に変換する共起行列変換ステップと、

前記複数の共起行列を入力として統計的特徴抽出方法により特徴ベクトルを求めるための基礎となる固有共起行列群を決定する固有共起行列群決定ステップと、

前記1以上のカテゴリを含む1以上のプロファイル学習用時系列データに対して前記ウィンドウ・データ取出ステップ、前記スコープ・データ抽出ステップ及び前記共起行列変換ステップと同様のステップをそれぞれ実施して、前記1以上のプロファイル学習用時系列データを1以上のプロファイル用共起行列に変換するプロファイル用共起行列変換ステップと、

前記1以上のプロファイル用共起行列と前記固有共起行列群とに基づいて前記1以上のプロファイル学習用時系列データについての1以上の判定用特徴ベクトルを抽出する判定用特徴ベクトル抽出ステップと、

テストの対象となるテスト時系列データに対して前記ウィンドウ・データ取出ステップ、前記スコープ・データ抽出ステップ及び前記共起行列変換ステップと同様のステップを実施して、前記テスト時系列データをテスト用共起行列に変換するテスト用共起行列変換ステップと、

前記テスト用共起行列と前記固有共起行列群とに基づいて前記テスト用時系列データについてのテスト用特徴ベクトルを抽出するテスト用特徴ベクトル抽出ステップと、

前記判定用特徴ベクトルと前記固有共起行列群を構成する複数の固有共起行列との掛け算に基づいて、前記複数の固有共起行列の次元を小さくした複数の判定用近似共起行列を取得する判定用近似共起行列取得ステップと、

前記複数の判定用近似共起行列からイベント抽出を行って判定用多層ネットワークモデルを生成する判定用多層ネットワークモデル生成ステップと、

前記テスト用特徴ベクトルと前記固有共起行列群を構成する複数の固有共起行列とを掛け算することにより、前記複数の固有共起行列の次元を小さくした複数のテスト用近似共起行列を取得するテスト用近似共起行列取得ステップと、

前記複数のテスト用近似共起行列からイベント抽出を行ってテスト用多層ネットワークモデルを生成するテスト用多層ネットワークモデル生成ステップと、

前記判定用多層ネットワークモデルと前記テスト用多層ネットワークモデルとに基づいて、前記テスト時系列データが前記1以上のカテゴリに属するか否かを判定する判定ステップとからなる時系列データ判定方法。

【請求項2】

前記判定ステップでは、前記判定用多層ネットワークモデルと前記テスト用多層ネットワークモデルとの類似度に基づいて前記テスト時系列データが前記1以上のカテゴリに属

するか否かを判定することを特徴とする請求項 1 に記載の時系列データ判定方法。

【請求項 3】

前記判定ステップでは、前記類似度が予め定めた閾値より大きいか否かにより、前記テスト時系列データが前記 1 以上のカテゴリに属するか否かを判定することを特徴とする請求項 2 に記載の時系列データ判定方法。

【請求項 4】

前記判定用多層ネットワークモデル及び前記テスト用多層ネットワークモデルは、それぞれ共起性が正の値から構成されるネットワークモデルと、共起性が負の値から構成されるネットワークモデルとからなり、

前記判定ステップでは、前記共起性が正の値から構成されるネットワークモデルと、前記共起性が負の値から構成されるネットワークモデルの少なくとも一方を用いて、前記類似度を求めることを特徴とする請求項 2 または 3 に記載の時系列データ判定方法。

【請求項 5】

請求項 1 乃至 4 のいずれか 1 項に記載の時系列データ判定方法を用いて、コンピュータシステムに入力される時系列データの異常を判別することを特徴とする時系列データ異常判別方法。