(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第6508353号
(P6508353)

(45)発行日　令和1年5月8日(2019.5.8)　(24)登録日　平成31年4月12日(2019.4.12)

(51)Int.Cl.		F I		
G06F 21/55	(2013.01)	G06F	21/55	320
G06F 21/62	(2013.01)	G06F	21/62	309
G06N 20/00	(2019.01)	G06N	99/00	153

請求項の数 9　(全 29 頁)

(21)出願番号　特願2017-545169(P2017-545169)
(86)(22)出願日　平成28年10月5日(2016.10.5)
(86)国際出願番号　PCT/JP2016/079637
(87)国際公開番号　WO2017/065070
(87)国際公開日　平成29年4月20日(2017.4.20)
審査請求日　平成30年4月4日(2018.4.4)
(31)優先権主張番号　特願2015-202280(P2015-202280)
(32)優先日　平成27年10月13日(2015.10.13)
(33)優先権主張国　日本国(JP)

(73)特許権者　000004237
日本電気株式会社
東京都港区芝五丁目７番１号
(74)代理人　100103090
弁理士　岩壁　冬樹
(74)代理人　100124501
弁理士　塩川　誠人
(72)発明者　友永　康之
東京都港区芝五丁目７番１号　日本電気株式会社内

審査官　上島　拓也

最終頁に続く

(54)【発明の名称】情報処理装置

(57)【特許請求の範囲】
【請求項１】
データに対する利用者の行動であるデータアクセス行動に関するアクセス情報であって、データにアクセスする利用者に由来する第１の情報と、アクセスされるデータに由来する第２の情報とを含むアクセス情報と、不審行動または正常行動との関係を示すアクセス行動モデルを記憶するモデル記憶手段と、
前記アクセス行動モデルに基づいて、任意のデータアクセス行動が不審行動であるか否かを判定する判定手段と、
前記アクセス行動モデルに基づいて、不審行動に該当するアクセス行動が行われる危険性があるデータを予測する危険データ予測手段とを備えた
ことを特徴とする情報処理装置。
【請求項２】
アクセス情報は、第１の情報として、アクセスする利用者、アクセスされる時間、アクセス種別もしくはアクセス方法に関する情報を含む、または、第２の情報として、データ自体もしくはデータの格納場所に関する情報を含む
請求項１に記載の情報処理装置。
【請求項３】
アクセス情報は、アクセスする利用者に関する情報として、当該利用者が生成したテキストに関する情報もしくは当該利用者が所定のデータに対して行ったアクセス行動に関する統計値を含む、または、データ自体に関する情報として、当該データの内容に関する情

報もしくは当該データに対して行われたアクセス行動に関する統計値を含む
請求項２に記載の情報処理装置。
【請求項４】
アクセス情報と、前記アクセス情報が示すデータアクセス行動が不審行動であるか否かを示す情報とを学習データに用いて、機械学習によりアクセス行動モデルを生成する学習手段を備えた
請求項１から請求項３のうちのいずれか１項に記載の情報処理装置。
【請求項５】
ファイルサーバによって管理されているファイルを、対象データとする情報処理装置であって、
モデル記憶手段は、所定のファイルに対するアクセス履歴に含まれるアクセス行動のうち指定された期間におけるアクセス行動に関するアクセス情報と、前記アクセス行動が不審行動か否かを判別可能な情報とを用いて機械学習されたアクセス行動モデルを記憶する
請求項１から請求項４のうちのいずれか１項に記載の情報処理装置。
【請求項６】
アクセス情報から、各々が多次元の数値からなる２以上の数値ベクトルを生成する数値ベクトル生成手段を備え、
モデル記憶手段は、前記２以上の数値ベクトルの組と、不審行動または正常行動との関係を示すアクセス行動モデルとの関係を示すアクセス行動モデルを記憶し、
判定手段は、前記アクセス行動モデルを用いて算出される、指定されたアクセス情報から生成される２以上の数値ベクトルの組に対する不審行動または正常行動の確度に基づいて、前記アクセス情報によって示されるデータアクセス行動が不審行動であるか否かを判定する
請求項１から請求項５のうちのいずれか１項に記載の情報処理装置。
【請求項７】
数値ベクトル生成手段として、
アクセス情報に含まれる第１情報から、多次元の数値からなる第１数値ベクトルを生成する第１数値ベクトル生成手段と、
アクセス情報に含まれる第２情報から、多次元の数値からなる第２数値ベクトルを生成する第２数値ベクトル生成手段とを備え、
モデル記憶手段は、前記第１数値ベクトルと前記第２数値ベクトルとの組と、不審行動または正常行動との関係を示すアクセス行動モデルを記憶し、
判定手段は、前記アクセス行動モデルを用いて算出される、指定されたアクセス情報に含まれる第１情報および第２情報から生成される前記第１数値ベクトルと前記第２数値ベクトルの組に対する不審行動または正常行動の確度に基づいて、前記アクセス情報によって示されるデータアクセス行動が不審行動であるか否かを判定する
請求項６に記載の情報処理装置。
【請求項８】
アクセス行動モデルに基づいて、不審行動に該当するデータアクセス行動を行う危険性がある利用者を予測する危険利用者予測手段を備えた
請求項１から請求項７のうちのいずれか１項に記載の情報処理装置。
【請求項９】
判定手段による判定結果に基づいて、アクセス権限を変更するアクセス権限変更手段を備えた
請求項１から請求項８のうちのいずれか１項に記載の情報処理装置。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、不審行動を検知するための不審行動検知システムに用いられる情報処理装置に関する。

【背景技術】
【０００２】
近年、企業データの情報漏えい対策が特に注目されている。特に、データに対する有効なアクセス権限を有する関係者が契機となって生じる情報漏えいに対する対策が注目されている。
【０００３】
企業データの情報漏えい事例の分析により、企業データに対して有効なアクセス権限を有する社内関係者や当該企業の外部委託先の担当者が契機となっているケースが多いことがわかってきたためである。
【０００４】
情報漏えい対策の代表的な例としては、全てのデータを暗号化する手法や、ルールベースで利用者の不審行動を検知して禁止する手法や、統計ベースで利用者の不審行動を検知して禁止する手法が挙げられる。なお、本発明では、データに対して正当な権限を有する利用者が権限を悪用して当該データにアクセスする行為を不審行動と呼ぶ。また、以下では、データに対して正当な権限を有する利用者が当該権限を正当に利用して（権限を設定した目的の範囲内で）当該データにアクセスする行為を正常行動と呼ぶ場合がある。この場合、あるデータに対して正当な権限を有する利用者の当該データに対するアクセス行動は、正常行動か不審行動のいずれかに分類される。
【０００５】
例えば、特許文献１には、上記の統計ベースで利用者の不審行動を検知する手法の例が記載されている。より具体的には、特許文献１に記載のシステムは、ユーザの操作ログから、所定の時間帯における所定の操作について、ユーザ毎に、操作状況の推移を演算する。そして、演算された操作状況の推移を示す数値から構成されるモデルを生成し、それらの平均値を求める。そして、各ユーザの操作状況の推移を示す数値と平均値との乖離計算により、特異な操作を行った利用者を検知する。
【０００６】
また、データから特徴量を得る技術に関連して、非特許文献１には、数値のみからなる多次元ベクトルに対して特徴抽出を行って、特徴ベクトルを生成する方法が記載されている。
【先行技術文献】
【特許文献】
【０００７】
【特許文献１】特開２００８－１９２０９１号公報
【非特許文献】
【０００８】
【非特許文献１】Bespalov, Dmitriy and Qi, Yanjun and Bai, Bing and Shokoufandeh, Ali、"Sentiment Classification with Supervised Sequence Embedding"、Machine Learning and Knowledge Discovery in Databases、vol.7523、2012、p.159-174
【発明の概要】
【発明が解決しようとする課題】
【０００９】
上記の全てのデータを暗号化する手法は、利用者がデータをそのまま持ち出しても専用ソフトウェアを使用しないと暗号化を解除できないため、情報漏えい対策として有効である。しかし、この手法は、通常の業務等で取引先企業にデータを送付する際に、都度当該データの暗号化を解除する権限を有する特権管理者に暗号化の解除を依頼する必要があり、生産性が低下する問題がある。また、この手法は、特定ファイルを暗号化対象から除外するなど抜け穴が生じる問題がある。また、この手法は、特権管理者がその権限を悪用してデータの暗号化を解除するケースを防止できない問題がある。
【００１０】
アクセスログ等を分析し、アクセスパターンに関するルールを設定して不審行動を検知

するなどのルールベースの手法は、特権管理者を含む全ての利用者に適用できるため、特権管理者の権限悪用による情報漏えいを防止できる可能性が高い。しかし、この手法は、予めルールを設定することが非常に困難であるという問題がある。また、この手法は、設定したルールのメンテナンスに手間がかかる等の問題がある。

【００１１】

なお、統計ベースの手法としては、特許文献１に記載されているような、利用者の通常行動と相関のある特徴量（例えば、１分間のファイルサーバアクセス数など）を計算し、この特徴量が予め設定した閾値を超過した場合に不審行動を検知する手法がある。しかし、特許文献１に記載の手法は、利用者の不審行動または通常行動と相関のある特徴量を決めるためにアクセスログを統計分析する必要があり、導入時の敷居が高いという問題がある。また、アクセスログの統計分析の対象とされる利用者およびデータに関する情報は、大量かつ多様なテキストを含むことが多い。この場合、特許文献１に記載の手法では、上記の特徴量が高次元となるが、統計分析でこのような高次元の特徴量を扱うのは困難である。このため、特許文献１に記載の手法は、不審行動の検知精度が低い問題がある。

【００１２】

そこで、本願発明は、上記に鑑み、予めルールを設定することなく、高精度に不審行動を検知できる不審行動検知システムに用いられる情報処理装置を提供することを目的とする。

【課題を解決するための手段】

【００１３】

本発明による情報処理装置は、データに対する利用者の行動であるデータアクセス行動に関するアクセス情報であって、データにアクセスする利用者に由来する第１の情報と、アクセスされるデータに由来する第２の情報とを含むアクセス情報と、不審行動または正常行動との関係を示すアクセス行動モデルを記憶するモデル記憶手段と、アクセス行動モデルに基づいて、任意のデータアクセス行動が不審行動であるか否かを判定する判定手段と、アクセス行動モデルに基づいて、不審行動に該当するアクセス行動が行われる危険性があるデータを予測する危険データ予測手段とを備えたことを特徴とする。

【発明の効果】

【００１７】

本発明によれば、予めルールを設定することなく、精度良く不審行動を検知できる。

【図面の簡単な説明】

【００１８】

【図１】第１の実施形態の不審行動検知システムの構成例を示すブロック図である。

【図２】第１の実施形態の不審行動検知システムの動作例を示すフローチャートである。

【図３】第１の実施形態の不審行動検知システムの他の構成例を示すブロック図である。

【図４】第１の実施形態の不審行動検知システムの他の動作例を示すフローチャートである。

【図５】第１の実施形態の不審行動検知システムの他の構成例を示すブロック図である。

【図６】数値ベクトル生成手段１６のより詳細な構成例を示すブロック図である。

【図７】第２の実施形態の不審行動検知システムの構成例を示すブロック図である。

【図８】利用者データ記憶部１０１が保持する利用者データのデータ構造の一例を示す説明図である。

【図９】文書データ記憶部１０２が保持する文書データのデータ構造の一例を示す説明図である。

【図１０】アクセスログ記憶部１０５が保持するアクセスログのデータ構造の一例を示す説明図である。

【図１１】予測スコア記憶部１１２が保持する予測結果のデータ構造の一例を示す説明図である。

【図１２】不審行動検知システム１００のアクセス行動学習ステップの動作例を示すフローチャートである。

【図１３】不審行動検知システム１００のアクセス行動予測ステップの動作例を示すフローチャートである。
【図１４】不審行動検知システム１００の不審行動通知ステップの動作例を示すフローチャートである。
【図１５】第２の実施形態の第１変形例の不審行動検知システムの構成例を示すブロック図である。
【図１６】第２の実施形態の第１変形例の不審行動検知システムの動作例を示すフローチャートである。
【図１７】第２の実施形態の第２変形例の不審行動検知システムの構成例を示すブロック図である。
【図１８】アクセス権限制御画面の例を示す説明図である。
【図１９】第２の実施形態の第２変形例の不審行動検知システムの動作例を示すフローチャートである。
【図２０】第２の実施形態の第３変形例の不審行動検知システムの構成例を示すブロック図である。
【図２１】第２の実施形態の第３変形例の不審行動検知システムの動作例を示すフローチャートである。
【発明を実施するための形態】
【００１９】
実施形態１．
以下、本発明の実施形態について図面を参照して説明する。図１は、本発明の第１の実施形態にかかる不審行動検知システムの構成例を示すブロック図である。図１に示す不審行動検知システム１０は、モデル記憶手段１１と、判定手段１２とを備える。
【００２０】
モデル記憶手段１１は、アクセス情報と不審行動との関係またはアクセス情報と正常行動との関係を示すアクセス行動モデルを記憶する。アクセス情報は、データに対する利用者の行動であるデータアクセス行動に関する情報であって、データにアクセスする利用者に由来する第１の情報と、アクセスされるデータに由来する第２の情報とを含む。
【００２１】
判定手段１２は、モデル記憶手段１１に記憶されているアクセス行動モデルに基づいて、任意のデータアクセス行動が不審行動であるか否かを判定する。
【００２２】
ここで、第１の情報は、例えば、データにアクセスする利用者に関する情報や、該利用者がデータにアクセスする際の時間（アクセス時間）、種別（アクセス種別）または方法（アクセス方法）に関する情報であってもよい。また、第２の情報は、アクセスされるデータ自体に関する情報（いわゆるデータの属性情報や、特徴量といったデータの内容に関する情報等）であってもよい。第２の情報は、データ自体に関する情報に限られず、例えば、該データの格納場所に関する情報や、該データに対して行われたアクセス行動に関する統計値であってもよい。
【００２３】
また、データにアクセスする利用者に関する情報は、一般に利用者の属性情報とされる情報に限られず、例えば、当該利用者が生成したテキストに関する情報や、当該利用者が所定のデータに対して行ったアクセス行動に関する統計値であってもよい。
【００２４】
また、図２は、本実施形態の動作例を示すフローチャートである。図２に示す例では、まず、判定手段１２は、モデル記憶手段１１からアクセス行動モデルを読み出す（ステップＳ１１）。次に、判定手段１２は、読み出されたアクセス行動モデルに基づいて、指定されたアクセス情報に対して、該アクセス情報が示すデータアクセス行動が不審行動であるか否かを判定する（ステップＳ１２）。
【００２５】

アクセス情報の取得方法としては、例えば、管理者が直接入力してもよいし、システムが、所定のデータに対するアクセス履歴に含まれる指定された期間やデータや利用者等の情報を基に生成してもよい。

【００２６】

このような構成によれば、データにアクセスした利用者に由来する情報と、アクセスされたデータに由来する情報という少なくとも２つの観点による情報の組からデータアクセス行動が不審行動か否かを判別可能なアクセス行動モデルを基に、任意のアクセス行動が不審行動か否かを判定できるため、予めルールを設定することなく、不審行動を高精度に検知できる。

【００２７】

また、図１に示す構成において、データは、ファイルサーバによって管理されているファイルであってもよい。そのような場合に、モデル記憶手段１１は、所定のファイルに対するアクセス履歴に含まれるアクセス行動のうち指定された期間におけるアクセス行動に関するアクセス情報と、該アクセス行動が不審行動か否かを判別可能な情報とを用いて機械学習されたアクセス行動モデルを記憶してもよい。

【００２８】

また、図３は、不審行動検知システム１０の他の構成例を示すブロック図である。図３に示すように、不審行動検知システム１０は、図１に示す構成に加えて、例えば、アクセス情報と、該アクセス情報が示すデータアクセス行動が不審行動か否かを判別可能な情報とを学習データに用いて、機械学習によりアクセス行動モデルを生成する学習手段１３を備えていてもよい。

【００２９】

そのような学習手段１３を備えることによって、学習手段に与えるデータの次元数が膨大であっても学習が可能になる。なお、データの次元数は、例えば、１０００以上であってもよいし、１００００以上であってもよい。

【００３０】

また、図３に示すように、不審行動検知システム１０は、例えば、判定手段１２による判定結果を基に、実際に行われたデータアクセス行動から不審行動を検知する不審行動検知手段１４をさらに備えていてもよい。

【００３１】

また、図４は、不審行動検知システム１０の図３に示す構成における動作例を示すフローチャートである。図４に示す例では、まず、学習手段１３が、アクセス情報と、該アクセス情報が示すデータアクセス行動が不審行動か否かを判別可能な情報とを学習データに用いて、機械学習によりアクセス行動モデルを生成する（ステップＳ２１）。また、学習手段１３は、生成したアクセス行動モデルをモデル記憶手段１１に書き込む（ステップＳ２２）。

【００３２】

次に、判定手段１２が、モデル記憶手段１１からアクセス行動モデルを読み出し、読み出されたアクセス行動モデルに基づいて、指定されたアクセス情報に対して不審行動か否かを判定する（ステップＳ１１、ステップＳ１２）。

【００３３】

判定手段１２による判定の結果、不審行動であった場合（ステップＳ２３のＹｅｓ）、不審行動検知手段１４は、指定されたアクセス情報が示すアクセス行動が不審行動であるとして、所定の検知処理を行う（ステップＳ２４）。検知処理は、例えば、検知した不審行動に関する情報を記憶したり、管理者に通知する処理であってもよい。

【００３４】

一方、不審行動でなかった場合（ステップＳ２３のＮｏ）、システムは、次のアクセス情報が指定されるのを待つ（ステップＳ１２に戻る）。

【００３５】

ステップＳ１２～ステップＳ２４までの動作を、例えば、アクセス情報が指定される度

に繰り返す。

【００３６】

また、図５は、不審行動検知システム１０の他の構成例を示すブロック図である。図５に示すように、不審行動検知システム１０は、例えば、通知手段１５や、数値ベクトル生成手段１６や、危険利用者予測手段１７や、危険データ予測手段１８や、アクセス権限変更手段１９をさらに備えていてもよい。

【００３７】

通知手段１５は、不審行動が検知されると、管理者に通知を行う。

【００３８】

数値ベクトル生成手段１６は、アクセス情報から、各々が多次元の数値からなる２以上の数値ベクトルを生成する。

【００３９】

数値ベクトル生成手段１６を備える構成において、モデル記憶手段１１は、数値ベクトル生成手段１６が生成した数値ベクトルの組と、不審行動または正常行動との関係を示すアクセス行動モデルとの関係を示すアクセス行動モデルを記憶してもよい。また、判定手段１２は、そのようなアクセス行動モデルを用いて算出される、指定されたアクセス情報から生成される２以上の数値ベクトルの組に対する不審行動または正常行動の確度に基づいて、当該アクセス情報によって示されるデータアクセス行動が不審行動であるか否かを判定してもよい。

【００４０】

また、図６は、数値ベクトル生成手段１６のより詳細な構成例を示すブロック図である。図６に示すように、数値ベクトル生成手段１６は、第１数値ベクトル生成手段１６１と、第２数値ベクトル生成手段１６２とを含んでいてもよい。

【００４１】

第１数値ベクトル生成手段１６１は、アクセス情報に含まれる第１情報から、多次元の数値からなる第１数値ベクトルを生成する。

【００４２】

第２数値ベクトル生成手段１６２は、アクセス情報に含まれる第２情報から、多次元の数値からなる第２数値ベクトルを生成する。

【００４３】

第１数値ベクトル生成手段１６１および第２数値ベクトル生成手段１６２を備える構成において、モデル記憶手段１１は、第１数値ベクトルと第２数値ベクトルの組と、不審行動または正常行動との関係を示すアクセス行動モデルを記憶してもよい。また、判定手段１２は、そのようなアクセス行動モデルを用いて算出される、指定されたアクセス情報から生成される第１数値ベクトルと第２数値ベクトルの組に対する不審行動または正常行動の確度に基づいて、当該アクセス情報によって示されるデータアクセス行動が不審行動であるか否かを判定してもよい。

【００４４】

危険利用者予測手段１７は、アクセス行動モデルに基づいて、データに対して、不審行動に該当するデータアクセス行動を行う危険性がある利用者を予測する。

【００４５】

危険データ予測手段１８は、アクセス行動モデルに基づいて、利用者に対して、不審行動に該当するアクセス行動が行われる危険性があるデータを予測する。

【００４６】

アクセス権限変更手段１９は、判定手段１２による判定結果、不審行動検知手段１４による検知結果、危険データ予測手段１８による予測結果または危険利用者予測手段１７による予測結果に基づいて、アクセス権限を変更する。

【００４７】

このような構成によれば、不審行動を高精度に検知できるだけでなく、検知した不審行動の情報（検知の対象となったアクセス情報など）を管理者に通知できる。また、不審行

動が検知された利用者（不審行動者）が、当該不審行動が検知されたデータ（対象データ）を不正に取得できないよう、当該利用者に対する対象データのアクセス権限を自動的に変更できる。また、事前に、そのような不審行動を行う可能性のある利用者や対象データを予測できるので、不審行動を未然に防ぐことができる。また、データのアクセス権限の設定に穴があってもその穴を塞ぐことができる。

【００４８】

本実施形態において、モデル記憶手段１１は、例えば、記憶装置によって実現される。また、判定手段１２、学習手段１３、不審行動検知手段１４、通知手段１５、数値ベクトル生成手段１６、危険利用者予測手段１７、危険データ予測手段１８およびアクセス権限変更手段１９は、例えば、プログラムに従って動作する情報処理装置によって実現される。なお、通知手段１５が、ディスプレイ装置等を介して管理者に情報の通知を行う場合、通知手段１５は、例えば、プログラムに従って動作する情報処理装置と、ディスプレイなどの表示装置または該表示装置とのインタフェース部とによって実現されてもよい。

【００４９】

実施形態２．

次に、本発明の第２の実施形態について説明する。なお、以下では、不審行動の検知対象とするデータが、ファイルサーバによって管理されるファイルである場合を例に用いて説明を行うが、データはファイルサーバによって管理されるファイルに限定されない。例えば、データは、データベースシステム等に格納される任意の単位のデータであってもよい。

【００５０】

まず、本実施形態の特徴を簡単に説明する。本実施形態の不審行動検知システムは、（１）ファイルサーバの利用者データ、（２）ファイルサーバが格納する文書データ、（３）ファイルサーバのアクセスログ、の３つのデータを用いて、各ファイルサーバ利用者の通常時におけるファイルサーバに対するアクセス行動を機械学習（教師あり学習）でモデル化する。そして、各ファイルサーバ利用者の実際のファイルサーバに対するアクセス行動と上記のモデルで予測されるアクセス行動の乖離を常時監視することにより、乖離の大きいファイルサーバ利用者を不審行動者として自動検知する。

【００５１】

ここで、（１）利用者データは、例えば、氏名、年齢、性別、学歴、担当業務、役職、部署、管理スパン（スパンオブコントロール）、異動履歴、保有資格、職務経歴、業績評価、健康診断結果などを含んでいてもよい。また、（２）文書データは、例えば、文書名、ファイルパス、アクセス権限、更新日時などのプロパティ設定、文書の内容に関する情報（テキスト、画像など）などを含んでいてもよい。また、（３）アクセスログは、ファイルサーバに対するアクセス履歴を保存したファイルであってもよい。なお、いずれのデータにおいても、大量かつ多様なテキストデータ（非構造化データ）が含まれていてもよい。

【００５２】

また、本実施形態の不審行動検知システムが行う不審行動検知方法には、前処理ステップ、特徴抽出ステップ、学習ステップ、予測ステップ、通知ステップ、の５つのプロセスを含まれる。

【００５３】

前処理ステップでは、上記の３つのデータ（利用者データ、文書データ、アクセスログ）から＜利用者属性、文書属性、アクセス実績＞のデータ組（tuple）を生成する。ここで、利用者属性は、ファイルサーバの利用者データから、利用者の特徴を表現するデータ項目の内容を抜き出したものであればよい。文書属性は、ファイルサーバが格納する文書データから、文書の特徴を表現するデータ項目の内容を抜き出したものであればよい。アクセス実績は、ファイルサーバのアクセスログで示される、当該利用者が当該文書にアクセスした実績の有無を判別可能な情報であればよい。例えば、アクセス実績は、アクセスした実績がある場合は１、ない場合は０等として２値化された情報であってもよい。

【００５４】
特徴抽出ステップでは、上記のデータ組のうち利用者属性および文書属性からそれぞれ特徴ベクトルを生成する。
【００５５】
学習ステップでは、上記のデータ組の集合から学習対象期間に該当するデータ組を切り出した上で、それらデータ組を用いて要素間の関係性（より具体的には、＜利用者属性、文書属性＞ペアとアクセス実績の関係性）を機械学習して予測モデルを生成する。機械学習アルゴリズムには、米国特許第８３４１０９５号明細書に記載された方法（Supervised Semantic Indexing（以下、ＳＳＩという））を用いることを想定するが、その他、一般的な機械学習手法を組み合わせてもよい。
【００５６】
予測ステップでは、上記のデータ組の集合から予測対象期間に該当するデータ組を切り出した上で、それらデータ組に対して予測モデルを適用する。より具体的には、それらデータ組の各々が示す＜利用者属性、文書属性＞ペアに対してアクセス行動の予測スコアを計算する。本実施形態では、予測スコアを［０．０～１．０］の実数値とする。なお、予測スコアが１．０に近いほど、当該＜利用者属性、文書属性＞ペアはアクセス確度が高いすなわち正常行動である可能性が高いことを表す。一方、予測スコアが０．０に近いほど、当該＜利用者属性、文書属性＞ペアはアクセス確度が低いすなわち不審行動である可能性が高いことを表す。
【００５７】
通知ステップでは、予測ステップで計算した＜利用者属性、文書属性＞ペアのうち、予測スコアが閾値（例えば０．１など）より低いもの（つまり当該利用者属性が示す利用者が当該文書属性が示す文書にアクセスする確度が低いと予測されるもの）を不審行動として抽出する。そして、抽出された不審行動の対象とされた利用者のリストを管理者等に通知する。
【００５８】
以下、より具体的な構成について説明する。図７は、本実施形態の不審行動検知システムの構成例を示すブロック図である。
【００５９】
図７に示す不審行動検知システム１００は、利用者データ記憶部１０１と、文書データ記憶部１０２と、利用者データ前処理部１０３と、文書データ前処理部１０４と、アクセスログ記憶部１０５と、アクセスログ前処理部１０６と、利用者属性特徴抽出部１０７と、文書属性特徴抽出部１０８と、アクセス実績学習部１０９と、予測モデル記憶部１１０と、予測スコア算出部１１１と、予測スコア記憶部１１２と、不審行動通知部１１３とを備える。
【００６０】
不審行動検知システム１００は、例えば、パーソナルコンピュータやサーバ装置等の情報処理装置と、該情報処理装置がアクセス可能なデータベースシステム等の記憶装置群とによって実現される。このとき、利用者データ前処理部１０３、文書データ前処理部１０４、アクセスログ前処理部１０６、利用者属性特徴抽出部１０７、文書属性特徴抽出部１０８、アクセス実績学習部１０９、予測スコア算出部１１１および不審行動通知部１１３は、例えば、情報処理装置が備えるＣＰＵによって実現されてもよい。その場合、該ＣＰＵは、所定の記憶装置に記憶された各処理部の動作を記述したプログラムを読み出し、該プログラムに従って動作することによって各処理部の機能を実現する。また、利用者データ記憶部１０１、文書データ記憶部１０２、アクセスログ記憶部１０５、予測モデル記憶部１１０および予測スコア記憶部１１２は、例えば、情報処理装置がアクセス可能な記憶装置群によって実現されてもよい。なお、記憶装置は１つであっても複数であってもよい。
【００６１】
利用者データ記憶部１０１は、ファイルサーバの利用者の利用者データを保持する。フ

ァイルサーバの利用者データの項目例としては、氏名、年齢、性別、学歴、担当業務、役職、部署、管理スパン、異動履歴、保有資格、職務経歴、業績評価、健康診断結果などが挙げられる。

【００６２】

図８は、利用者データ記憶部１０１が保持する利用者データのデータ構造の一例を示す説明図である。図８に示すように、利用者データ記憶部１０１は、利用者データとして、例えば、利用者を識別する利用者ＩＤと対応づけて、利用者の氏名、年齢、性別、役職、担当業務、業績評価等の情報を記憶してもよい。利用者データは、さらに利用者の人物像や勤務態度に関する説明等がテキスト形式で記載された情報を含んでいてもよい。また、利用者データは、さらに健康診断結果を含んでいてもよい。なお、図８において、網掛けは一人分の利用者データに相当するレコードの例を示している。

【００６３】

文書データ記憶部１０２は、ファイルサーバが格納する文書の文書データを保持する。文書データの項目例としては、文書名、文書の種別、ファイルパス、アクセス権限、更新日時などの当該文書に付随するプロパティ設定などが挙げられる。

【００６４】

図９は、文書データ記憶部１０２が保持する文書データのデータ構造の一例を示す説明図である。図９に示すように、文書データ記憶部１０２は、文書データとして、例えば、文書を識別する文書ＩＤと対応づけて、文書の種別、アクセス権限の設定内容、作成日時、更新日時等のプロパティ情報を記憶してもよい。また、文書データは、さらに文書の内容に関する説明等がテキスト形式で記載された情報を含んでいてもよい。なお、図９において、網掛けは一ファイル分の文書データに相当するレコードの例を示している。

【００６５】

利用者データ前処理部１０３は、利用者データ記憶部１０１を参照して、指定された利用者に関するレコードを読み込む。また、利用者データ前処理部１０３は、読み込んだレコードに含まれる指定された利用者に関する情報（以下、利用者属性情報という場合がある）を用いて、利用者ベクトルを生成する。ここで、利用者ベクトルは、利用者属性情報が示す内容を、数値からなる多次元ベクトルで表現したものである。利用者データ前処理部１０３は、例えば、上記処理を利用者属性特徴抽出部１０７の命令に応じて行う。

【００６６】

文書データ前処理部１０４は、文書データ記憶部１０２を参照して、指定された文書に関するレコードを読み込む。また、文書データ前処理部１０４は、読み込んだレコードに含まれる指定された文書に関する情報（以下、文書属性情報という場合がある）を用いて、文書ベクトルを生成する。ここで、文書ベクトルは、文書属性情報が示す内容を、数値からなる多次元ベクトルで表現したものである。文書データ前処理部１０４は、例えば、上記処理を文書属性特徴抽出部１０８の命令に応じて行う。

【００６７】

アクセスログ記憶部１０５は、所定のファイルサーバのアクセスログを保持する。ファイルサーバのアクセスログには、ファイルサーバ利用者がファイルサーバにアクセスする度に、アクセス日時、アクセス者、アクセス文書等のアクセス行動に関する情報が記録される。

【００６８】

図１０は、アクセスログ記憶部１０５が保持するアクセスログのデータ構造の一例を示す説明図である。

【００６９】

アクセスログ前処理部１０６は、アクセスログ記憶部１０５を参照して、指定期間のアクセス日時をもつレコードを読み込む。また、アクセスログ前処理部１０６は、読み込んだレコードに含まれるアクセス者ＩＤおよびアクセス文書ＩＤを基に、ラベル情報を生成する。例えば、アクセスログ前処理部１０６は、アクセスログの指定期間中のレコードに含まれているアクセス者ＩＤとアクセス文書ＩＤの組を用いて、該アクセス者ＩＤに対応

する利用者ＩＤと該アクセス文書ＩＤに対応する文書ＩＤの組に対して、正否ラベルを正解（１）とするラベル情報＜利用者ＩＤ、文書ＩＤ、正否ラベル（０／１）＞を生成してもよい。また、アクセスログ前処理部１０６は、例えば、アクセスログの指定期間中にアクセス実績のない利用者と文書の組をランダムに選択し、その利用者の利用者ＩＤとその文書の文書ＩＤの組に対して、正否ラベルを不正解（０）とするラベル情報を生成してもよい。なお、アクセスログ前処理部１０６は、正解ラベル情報として、正常行動を行った利用者と文書の組を示すラベル情報＜利用者ＩＤ、文書ＩＤ＞を生成したり、不正解ラベル情報として、不審行動を行った利用者と文書の組を示すラベル情報＜利用者ＩＤ、文書ＩＤ＞を生成してもよい。以下、正解ラベル情報と不正解ラベル情報とを特に区別せず、不審行動か否かを判別可能なラベル情報という意味で正否ラベル情報と呼ぶ場合がある。アクセスログ前処理部１０６は、例えば、上記処理をアクセス実績学習部１０９の命令に応じて行う。

【００７０】

利用者属性特徴抽出部１０７は、利用者データ前処理部１０３が生成した利用者ベクトルに対して特徴抽出を行い、利用者特徴ベクトルを生成する。ここで、利用者特徴ベクトルは、利用者ベクトルの次元数よりも少ない次元数の数値ベクトルであればよい。利用者属性特徴抽出部１０７は、例えば、上記処理をアクセス実績学習部１０９または予測スコア算出部１１１の命令に応じて行う。

【００７１】

文書属性特徴抽出部１０８は、文書データ前処理部１０４が生成した文書ベクトルに対して特徴抽出を行い、文書特徴ベクトルを生成する。ここで、文書特徴ベクトルは、文書ベクトルの次元数よりも少ない次元数の数値ベクトルであればよい。文書属性特徴抽出部１０８は、例えば、上記処理をアクセス実績学習部１０９または予測スコア算出部１１１の命令に応じて行う。

【００７２】

アクセス実績学習部１０９は、利用者属性特徴抽出部１０７が生成した利用者特徴ベクトルと、文書属性特徴抽出部１０８が生成した文書特徴ベクトルと、アクセスログ前処理部１０６が生成したラベル情報とから、学習データとして＜利用者特徴ベクトル、文書特徴ベクトル、正否ラベル（１／０）＞を生成する。なお、ラベル情報は、正否ラベルを含むラベル情報（＜利用者ＩＤ、文書ＩＤ、正否ラベル＞）であっても、正否ラベルを含まない正否ラベル情報（＜利用者ＩＤ、文書ＩＤ＞）であってもよい。また、アクセス実績学習部１０９は、生成された学習データを用いて、利用者特徴ベクトルと文書特徴ベクトルと正否ラベルの間の関係性を機械学習し、予測モデルを生成する。

【００７３】

予測モデル記憶部１１０は、アクセス実績学習部１０９が生成した予測モデルを保持する。

【００７４】

予測スコア算出部１１１は、指定された利用者と文書のペアについて、予測データ＜利用者特徴ベクトル、文書特徴ベクトル＞を生成する。また、予測スコア算出部１１１は、生成した予測データに予測モデル記憶部１１０が保持する予測モデルを適用して、当該予測データに対するアクセス行動の予測スコアを算出する。予測スコア算出部１１１は、例えば、利用者と文書を指定して、利用者データ前処理部１０３、利用者属性特徴抽出部１０７、文書データ前処理部１０４および文書属性特徴抽出部１０８に命令をすることにより、予測データの要素を生成してもよい。

【００７５】

予測スコア記憶部１１２は、予測スコア算出部１１１による予測結果（予測スコアの算出結果）を、予測に用いた利用者と文書の情報とともに保持する。

【００７６】

図１１は、予測スコア記憶部１１２が保持する予測結果のデータ構造の一例を示す説明図である。図１１に示すように、予測スコア記憶部１１２は、例えば、アクセスする利用

者を識別するアクセス者ＩＤと、アクセスされたデータを識別されるアクセス文書ＩＤとともに、算出された予測スコアを記憶してもよい。

【００７７】

不審行動通知部１１３は、予測スコア記憶部１１２を参照し、予測スコアが閾値（例えば０．１など）より低いレコード（つまりアクセス確度が低いと予測されるレコード）を不審行動として抽出する。また、不審行動通知部１１３は、抽出した不審行動の対象とされた利用者のリストを、管理者等に所定の方法を用いて通知する。

【００７８】

次に、本実施形態の動作について説明する。本実施形態の不審行動検知システム１００の動作は、大きく、アクセス行動学習ステップ、アクセス行動予測ステップ、不審行動通知ステップ、の３つのステップに分類される。

【００７９】

アクセス行動学習ステップでは、アクセス実績学習部１０９が、利用者属性特徴抽出部１０７が生成した利用者特徴ベクトルと、文書属性特徴抽出部１０８が生成した文書特徴ベクトルと、アクセスログ前処理部１０６が生成したラベル情報とに基づいて、学習データを生成し、学習データの要素間の関係性、より具体的には利用者特徴ベクトルと文書特徴ベクトルの組に対する成否ラベルの関係性を機械学習して予測モデルを生成する。また、アクセス実績学習部１０９は、生成した予測モデルを予測モデル記憶部１１０に書き込む。

【００８０】

行動予測ステップでは、予測スコア算出部１１１が、指定された利用者および文書について、利用者特徴ベクトルと文書特徴ベクトルの組に対して予測モデルを適用し、当該利用者が当該文書にアクセスする確度を予測スコアとして計算する。また、予測スコア算出部１１１は、算出した予測スコアを、算出に用いた利用者と文書の情報とともに予測スコア記憶部１１２に書き込む。

【００８１】

不審行動通知ステップでは、不審行動通知部１１３が、予測スコア記憶部１１２から、予測スコアが閾値より低いレコードを不審行動として抽出し、抽出された不審行動に関する情報のリストを出力する。

【００８２】

図１２は、不審行動検知システム１００のアクセス行動学習ステップの動作例を示すフローチャートである。図１２に示す例では、まず、アクセス実績学習部１０９が、アクセスログ前処理部１０６を駆動し、アクセスログのうち指定した期間（つまり学習期間）のアクセス日時を有するレコードを読み出させる（ステップＳ１０１）。

【００８３】

ステップＳ１０１で、アクセスログ前処理部１０６は、例えば、アクセスログ記憶部１０５より、アクセス日時が条件にマッチするレコードをアクセス実績として読み込み、正解ラベル＜利用者ＩＤ、文書ＩＤ、正解ラベル（１）＞を生成してもよい。また、アクセスログ前処理部１０６は、例えば、読み込んだレコードに含まれる利用者ＩＤに対して、アクセス実績のない文書ＩＤをランダムに選択して、不正解ラベル＜利用者ＩＤ、文書ＩＤ、不正解ラベル（０）＞を生成してもよい。

【００８４】

次に、アクセス実績学習部１０９は、アクセス実績の件数分、ステップＳ１０３～ステップＳ１０８の動作を繰り返す（ステップＳ１０２、ステップＳ１０９）。

【００８５】

ステップＳ１０３では、アクセス実績学習部１０９が、利用者データ前処理部１０３を駆動し、ステップＳ１０１で読み出されたアクセス実績の利用者ＩＤの利用者データである利用者属性情報を読み出させる。また、利用者データ前処理部１０３は、読みだされたレコードの内容（利用者属性情報）をベクトル形式に変換し、利用者ベクトルを生成する。

【００８６】
利用者属性情報のベクトル化（数値化）は、例えば次のように実施される。すなわち、利用者データ前処理部１０３は、利用者属性情報のうち、年齢、年齢、最終学歴、資格などの値域が予め定まっている項目であるコード項目のデータであれば、予め定めたベクトル要素の値として当該コード項目の内容が所定の範囲に該当すれば１とし、該当しなければ０としてもよい（２値化）。
【００８７】
また、利用者データ前処理部１０３は、例えば、利用者属性情報のうち、テキスト形式の項目であるテキスト項目のデータであれば、当該テキスト項目の内容とされたテキストを形態素解析等を用いて単語に分解し、テキスト全体における単語または単語群の頻度等を計数してもよい。頻度は、一単語ごとではなく、二単語から五単語程度の単語群をひとまとめにして数えてもよい。最適な単語数は、学習対象となる利用者の数や文書の量によって異なる。また、利用者データ前処理部１０３は、例えば、計数された頻度を、当該単語または当該単語群に対応するベクトル要素の値としてもよい。
【００８８】
モデル学習ステップでは、後述する機械学習パラメータの更新時に、学習対象のデータ（文書特徴ベクトルと利用者特徴ベクトルの組）の一部を学習対象から外したデータでモデルを学習しなおして精度を検証する動作が行われる場合がある。そのとき、利用者データ前処理部１０３は、単語数を変えて検証することで、最適な単語数を定めてもよい。また、利用者データ前処理部１０３は、すべての文書において高頻度の単語、例えば助詞を除外するなど、頻度を数える対象となる単語を制限してもよい。そのようにして、テキストの特徴、つまりそのテキストを書いた利用者の特徴が表現された数値ベクトル（数値のみからなるデータ羅列）を生成する。
【００８９】
なお、利用者がＷｅｂサイトやＳＮＳに投稿したテキスト等も、利用者の特徴を表すデータ（数値）に変換可能である。昨今では、多くの人が興味を持った事柄について、ＳＮＳやブログ等に書き込んでいるため、それらの情報を用いることで、利用者の特徴を多く含む数値ベクトルを生成できる。
【００９０】
また、利用者データ前処理部１０３は、上記のテキストを数値化する方法と同様の方法で、アクセス先のＵＲＬ名等を分解し、それらに含まれる単語や単語群の頻度または滞留時間をカウントしたり、ＵＲＬ先のＨＴＴＰ文書を分解して、含まれる単語や単語群の頻度をカウントしてもよい。そのようなＷｅｂアクセス履歴に関する計数の結果も、ベクトル（数値）化できる。
【００９１】
ステップＳ１０４では、アクセス実績学習部１０９が、利用者属性特徴抽出部１０７を駆動し、ステップＳ１０３で生成された利用者ベクトルに対して特徴抽出を行わせて、利用者特徴ベクトルを生成させる。
【００９２】
一般に、ステップＳ１０３で生成される利用者ベクトルは、非常に大きいベクトル長のデータである。このため、そのままでは後段の学習および予測への適用が困難である。そこで、本実施形態では、利用者属性特徴抽出部１０７を用いて、利用者属性情報のうち特徴となるデータ項目のみを選択させ、データ長が圧縮されたベクトルを生成する。
【００９３】
利用者属性特徴抽出部１０７は、例えば、上述した非特許文献１に記載されている方法を利用して、特徴ベクトルを生成してもよい。なお、非特許文献１に記載された方法は全て自動で特徴ベクトルを生成するが、その方法以外にも、主成分分析などにより重要なベクトル項をまず手動で分析した上で、そのようなベクトル項を指定してもよい。そのような場合、利用者属性特徴抽出部１０７は、そのベクトル項の内容を表現した特徴ベクトルを生成してもよい。

【００９４】
ステップＳ１０５では、アクセス実績学習部１０９が、文書データ前処理部１０４を駆動して、ステップＳ１０１で読み込まれたアクセス実績の文書ＩＤの文書データ（文書属性情報）を読み出させる。文書データ前処理部１０４は、文書データ記憶部１０２より文書ＩＤがマッチするレコードを読み出して、ベクトル形式に変換して文書ベクトルを生成する。文書属性情報のベクトル化（数値化）は、ステップＳ１０３で示した利用者属性情報のベクトル化と同様の方法が適用可能である。
【００９５】
ステップＳ１０６では、アクセス実績学習部１０９が、文書属性特徴抽出部１０８を駆動し、ステップＳ１０５で生成された文書ベクトルに対して特徴抽出を行って、文書特徴ベクトルを生成させる。文書ベクトルからの特徴抽出は、ステップＳ１０４で示した利用者ベクトルからの特徴抽出方法と同様の方法が適用可能である。
【００９６】
ステップＳ１０７では、アクセス実績学習部１０９が、学習の前処理として、ステップＳ１０４で生成された利用者特徴ベクトルと、ステップＳ１０６で生成された文書特徴ベクトルのコサイン類似度を計算する。なお、本例では、２つのベクトルの類似度を測定するメトリックとしてコサイン類似度を用いているが、その他にも、任意のノルム（Ｌ１ノルム、Ｌ２ノルム、等）を用いることも可能である。
【００９７】
ステップＳ１０８では、アクセス実績学習部１０９が、ステップＳ１０７で計算された類似度と、ステップＳ１０１で生成されたラベル情報とを用いて、機械学習パラメータを調整する。
【００９８】
なお、本例では、機械学習の手段として、上述したＳＳＩを想定しているが、任意の教師あり機械学習分類器が適用可能である。任意の教師あり機械学習分類器の例として、サポートベクタマシン、ニューラルネット、ベイズ分類器などが広く知られている。
【００９９】
不審行動検知システムは、アクセス実績の件数分だけ上記の処理を繰り返すと、ステップＳ１１０に進む。
【０１００】
ステップＳ１１０では、アクセス実績学習部１０９が、ステップＳ１０８で調整された機械学習パラメータを予測モデル記憶部１１０に書き込む。
【０１０１】
また、図１３は、不審行動検知システム１００のアクセス行動予測ステップの動作例を示すフローチャートである。
【０１０２】
図１３に示す例では、まず、予測スコア算出部１１１が、ステップＳ１１０で書き込まれた調整済みの機械学習パラメータを予測モデル記憶部１１０から読み出す（ステップＳ２０１）。
【０１０３】
次に、予測スコア算出部１１１は、アクセスログ前処理部１０６を駆動し、アクセスログのうち指定した期間（予測期間）のアクセス日時を有するレコードを読み出させる（ステップＳ２０２）。ステップＳ２０２で、アクセスログ前処理部１０６は、読み出したレコード群を基に、ラベル情報＜利用者ＩＤ、文書ＩＤ、正否ラベル＞のリストを生成する。以下、ここで生成されたラベル情報のリストを、アクセス行動予測対象リストと呼ぶ場合がある。
【０１０４】
次に、予測スコア算出部１１１は、ステップＳ２０２で生成されたリストに含まれるレコードの件数分だけ、ステップＳ２０４～ステップＳ２０９の処理を繰り返す（ステップＳ２０３、ステップＳ２１０）。

【０１０５】
ステップＳ２０４では、予測スコア算出部１１１が、アクセス行動予測対象リストに含まれるラベル情報を順次取り出す。そして、予測スコア算出部１１１は、利用者データ前処理部１０３を駆動し、取り出したラベル情報に含まれる利用者ＩＤが示す利用者の利用者データを読み出させる。ステップＳ２０４で、利用者データ前処理部１０３は、指定された利用者ＩＤにマッチするレコード（利用者属性情報）を利用者データ記憶部１０１から読み出し、ベクトル形式に変換して利用者ベクトルを生成する。利用者属性情報のベクトル化（数値化）の方法は、ステップＳ１０３で示した方法と同様でよい。
【０１０６】
ステップＳ２０５では、予測スコア算出部１１１は、利用者属性特徴抽出部１０７を駆動して、ステップＳ２０４で生成された利用者ベクトルに対して特徴抽出を行わせて利用者特徴ベクトルを生成させる。利用者ベクトルの特徴抽出の方法は、ステップＳ１０４で示した方法と同様でよい。
【０１０７】
ステップＳ２０６では、予測スコア算出部１１１は、文書データ前処理部１０４を駆動し、ステップＳ２０４で取り出したラベル情報に含まれる文書ＩＤが示す文書の文書データを読み出させる。ステップＳ２０６で、文書データ前処理部１０４は、指定された文書ＩＤにマッチするレコード（文書属性情報）を文書データ記憶部１０２から読み出し、ベクトル形式に変換して文書ベクトルを生成する。文書属性情報のベクトル化（数値化）の方法は、ステップＳ１０３に示す方法と同様でよい。
【０１０８】
ステップＳ２０７では、予測スコア算出部１１１は、文書属性特徴抽出部１０８を駆動し、ステップＳ２０６で生成された文書ベクトルに対して特徴抽出を行わせて文書特徴ベクトルを生成させる。文書ベクトルの特徴抽出の方法は、ステップＳ１０４に示す方法と同様でよい。
【０１０９】
ステップＳ２０８では、予測スコア算出部１１１は、ステップＳ２０５で生成された利用者特徴ベクトルと、ステップＳ２０７で生成された文書特徴ベクトルとを用いて、ステップＳ２０１で読み出された機械学習パラメータに基づき、該利用者特徴ベクトルと該文書特徴ベクトルの組に対するアクセス確度を予測スコアとして算出する。既に説明したように、本例では、予測スコアを［０．０～１．０］の実数値とする。予測スコアは、例えば、サポートベクタマシンのprobability（確信度、信頼度）と呼ばれる数値であってもよい。
【０１１０】
ステップＳ２０９では、予測スコア算出部１１１が、ステップＳ２０８で算出された予測スコアと予測スコアの算出対象とされた利用者および文書の組とともに、予測結果を予測スコア記憶部１１２に書き込む。予測スコア算出部１１１は、＜利用者ＩＤ、文書ＩＤ、予測スコア＞の形式で予測結果を予測スコア記憶部１１２に書き込んでもよい。
【０１１１】
アクセス行動予測対象リストに含まれるレコードの件数分、上記処理を繰り返すと、当該行動予測ステップを終了する。
【０１１２】
また、図１４は、不審行動検知システム１００の不審行動通知ステップの動作例を示すフローチャートである。
【０１１３】
図１４に示す例では、まず、不審行動通知部１１３が、予測結果＜利用者ＩＤ、文書ＩＤ、予測スコア＞のリストである予測結果リストを読み出す（ステップＳ３０１）。
【０１１４】
次に、不審行動通知部１１３は、予測結果リストに含まれる予測結果の件数分、ステップＳ３０３～ステップＳ３０４の処理を繰り返す（ステップＳ３０２、ステップＳ３０５

）。
【０１１５】
ステップＳ３０３では、不審行動通知部１１３が、ステップＳ３０１で読み出されたレコードの予測スコアとあらかじめ設定した閾値（例えば０．１など）とを比較する。ここで、読み出したレコードの予測スコアが所定の閾値未満であれば、不審行動通知部１１３は、そのレコードが示す利用者と文書の組によるアクセス行動は不審行動であると判定する（ステップＳ３０３のＹｅｓ）。そして、不審行動通知部１１３は、ステップＳ３０４に進む。一方、所定の閾値以上であれば、不審行動通知部１１３は、その組によるアクセス行動は不審行動に該当しないすなわち正常行動であると判定する（ステップＳ３０３のＮｏ）。不審行動通知部１１３は、その後は特に処理は行わず、リストの次のレコードに処理を移すべくステップＳ３０３に戻る
【０１１６】
ステップＳ３０４では、不審行動通知部１１３が、不審行動とされた利用者と文書の組における少なくとも利用者の情報（利用者ＩＤ）を、一時記憶に記憶する。なお、不審行動通知部１１３は、利用者の情報だけでなく文書の情報（文書ＩＤ）や算出された予測スコアなども併せて記憶してもよい。このとき、不審行動通知部１１３は、繰り返し処理により同じ情報が登録済みの場合は、再度の登録をしなくてよい。
【０１１７】
予測結果リストの件数分、上記処理を完了すると、不審行動通知部１１３は、ステップＳ３０４で一時記憶に登録された情報を読み出し、不審行動として管理者等に通知する（ステップＳ３０６）。不審行動通知部１１３は、例えば、一時記憶に登録された情報に含まれる利用者ＩＤが示す利用者を、不審行動者として通知してもよい。また、不審行動通知部１１３は、例えば、一時記憶に登録された情報に含まれる文書ＩＤが示す文書を、通常時とは異なるアクセス行動が行われている危険文書として通知してもよい。
【０１１８】
以上のように、本実施形態では、データにアクセスする利用者の情報である利用者データ、データそのものの情報である文書データおよびアクセスログを用いて、不審行動の予測モデルを生成し、生成された予測モデルを基に、不審行動を検知している。このため、統計ベースで生成されるモデル等と比べて、扱えるデータ量を多く出来るので、より高精度な検知が可能となる。
【０１１９】
変形例１．
上記の実施形態では、検知した不審行動を通知するまでを実施する構成を示したが、不審行動検知システムは、不審行動が検知された利用者に対する対象データのアクセス権限の設定を自動で変更することも可能である。そのようにして、アクセス権限の穴を自動的に塞ぐことにより、ファイルサーバの利用者がデータを不正に持ち出す行為をプロアクティブに抑止することができる。
【０１２０】
図１５は、本変形例による不審行動検知システムの構成例を示すブロック図である。図１５に示す不審行動検知システム１００は、図７に示した構成に比べて、アクセス権限制御部１１４、アクセス権限記憶部１１５をさらに備えている点で異なる。
【０１２１】
アクセス権限制御部１１４は、不審行動の検知対象とされたデータを含む所定のデータに適用されるアクセス権限の設定、変更等の制御を行う。
【０１２２】
アクセス権限記憶部１１５は、不審行動の検知対象とされたデータを含む所定のデータに適用される現在のアクセス権限の情報を少なくとも保持する。
【０１２３】
図１６は、本変形例による不審行動検知システムの動作例を示すフローチャートである。本変形例では、上記の構成に比べて、さらに、アクセス権限制御ステップを含む。なお

、図１６は、本変形例による不審行動検知システム１００のアクセス権限制御ステップの動作例を示している。

【０１２４】

アクセス権限制御ステップでは、アクセス行動予測ステップによる予測スコアの算出結果に基づいて検知された不審行動の情報を基に、当該不審行動を行った利用者が同様のアクセス行動を行うことができなくなるようアクセス権限を制御する。アクセス権の制御は、例えば、検知された不審行動の対象とされたデータに対して、該不審行動が検知された利用者のアクセスを禁止するものであってもよい。例えば、アクセス権限制御部１１４が、不審行動の情報から利用者ＩＤと文書ＩＤとを取得し、該利用者ＩＤが示す利用者が該文書ＩＤが示す文書（データ）にアクセスできなくなるよう、該文書を格納するファイルサーバのホスト名を取得してアクセス権限を設定してもよい。

【０１２５】

図１６に示す例では、まず、アクセス権限制御部１１４が、不審行動通知部１１３より、検知された不審行動に関する情報を取得する（ステップＳ４０１）。

【０１２６】

次に、アクセス権限制御部１１４は、不審行動の対象文書を格納するファイルサーバのホスト名を取得する（ステップＳ４０２）。

【０１２７】

次に、アクセス権限制御部１１４は、不審行動者に対する当該ファイルサーバもしくは不審行動の対象文書のアクセス権限設定を変更する（ステップＳ４０３）。なお、アクセス権限設定の変更方法は、特に問わない。例えば、一般的に行われている方法を用いればよい。一例として、ディレクトリサービス（Windows（登録商標）の場合はActive DirectoryやLDAP）により管理される場合には、当該サービスを経由して、ファイルサーバ等のアクセス権限設定を変更する方法が挙げられる。

【０１２８】

変形例２．

また、第１変形例では、検知された不審行動に基づいて、アクセス権限設定の穴を自動的に塞ぐ例を示したが、システムは、運用担当者等の特定ユーザに、不審行動の情報とともに当該不審行動にかかるアクセス権限の設定変更を提案し、応答を待った上でアクセス権限の制御を行うことも可能である。そのようにすれば、実運用において、データやファイルサーバのアクセス権限設定が自動的に変更されてしまうことで、現場の業務が混乱することを防止できる。

【０１２９】

図１７は、本変形例の不審行動検知システムの構成例を示すブロック図である。図１６に示すよう不審行動検知システム１００は、図１５に示した構成に比べて、アクセス権限制御画面部１１６をさらに備えている点で異なる。

【０１３０】

アクセス権限制御画面部１１６は、後述するアクセス権限制御画面の制御を介して、特定ユーザに、特定ユーザに、不審行動にかかるアクセス権限の設定変更を行うか否かを問い合わせる。

【０１３１】

図１８は、アクセス権限制御画面の例を示す説明図である。図１８に示すように、アクセス権限制御画面は、ユーザに、不審行動者に対する当該ファイルサーバもしくは不審行動の対象文書のアクセス権限設定として、現状のアクセス許可設定を削除する（塞ぐ）か否（見逃す）かを選択させるものであってもよい。

【０１３２】

また、図１９は、本変形例による不審行動検知システムの動作例を示すフローチャートである。なお、図１９は、本変形例による不審行動検知システム１００のアクセス権限制御ステップの動作例を示している。

【０１３３】

図１９に示す例は、図１６に示した第１変形例における動作に、アクセス権限設定の制御を行うか否かの判定ステップ（ステップＳ５０１）が加わっている。

【０１３４】

例えば、ステップＳ５０１では、アクセス権限制御画面部１１６が、検知された不審行動者の利用者ＩＤと、当該不審行動者による不審行動の対象とされた文書を格納するファイルサーバのホスト名が少なくとも表示されており、かつ、「塞ぐ」および「見逃す」ボタン等、アクセス権限の制御を行うか否かを指示するＵＩ（ユーザインタフェース）部品を含むアクセス権限制御画面を表示してもよい。このとき、ファイルサーバの運用担当者等の特定ユーザは、画面の表示内容を確認した上で、当該人物が当該ファイルサーバにアクセスできなくなるようアクセス権限を制御するかどうかを判断すればよい。

【０１３５】

特定ユーザが［塞ぐ］ボタンを押下すると、アクセス権限制御画面部１１６は、ステップＳ４０３に進めばよい。一方、［見逃す］ボタンが押下されると、何も処理をせず終了してもよい。

【０１３６】

なお、不審行動が複数検知された場合は、その各々について上記の処理を行う。例えば、アクセス権限制御画面部１１６は、複数の不審行動の各々について、不審行動者の利用者ＩＤと、当該不審行動者による不審行動の対象とされた文書を格納するファイルサーバのホスト名が少なくとも表示されており、かつ、「塞ぐ」および「見逃す」ボタン等、アクセス権限の制御を行うか否かを指示するＵＩ（ユーザインタフェース）部品を含むアクセス権限制御画面を表示してもよい。

【０１３７】

なお、図１８に示す例では、不審行動者の利用者ＩＤと、当該不審行動者による不審行動の対象文書を格納するファイルサーバのホスト名の両方を表示しているが、いずれか一方の情報のみを表示してもよい。例えば、不審行動者の利用者ＩＤのみを取得・表示して、当該利用者ＩＤの利用者は不審行動を行う危険性があるとし、当該利用者に対するすべてのデータのアクセスを禁止するような、アクセス権限の設定を提案してもよい。また、例えば、不審行動の対象とされた文書を格納するファイルサーバのホスト名のみを取得・表示して、当該ファイルサーバもしくは当該文書は不審行動が行われる危険性があるとし、当該ファイルサーバに対するすべての利用者のアクセスを禁止するようなアクセス権限の設定を提案してもよい。

【０１３８】

なお、上記のアクセス権限の設定は、システムが自動設定する場合においても適用が可能である。

【０１３９】

変形例３．

また、本実施形態および各変形例では、アクセス行動学習ステップ、アクセス行動予測ステップ、不審行動通知ステップ、の３つのステップをすべて同一装置で実施する例を示したが、ネットワーク経由で（例えば、インターネット上に公開された予測モデルの配信サーバなどから）で予測モデルを受信する構成であれば、アクセス行動学習ステップを省略することも可能である。

【０１４０】

図２０は、本変形例の不審行動検知システムの構成例を示すブロック図である。図２０に示す構成は、図７に示す第１変形例の構成と比べて、上記のアクセス行動学習ステップでのみ用いられる要素（より具体的には、アクセスログ記憶部１０５、アクセスログ前処理部１０６およびアクセス実績学習部１０９）を省略し、新たに予測モデル受信部１１７が追加されている点が異なる。なお、これらの変更点を、例えば、他の変形例に適用することも可能である。

【０１４１】

予測モデル受信部１１７は、外部から予測モデルを受信する。予測モデルは、例えば、

当該システムを構成する装置以外の装置によって生成された予測モデルであってもよい。受信する予測モデルは、当該システムが不審行動の検出対象とするデータに対するアクセス行動に基づいて学習されたものでなくてもよい。例えば、稼動実績が十分あったり、アクセス権限等による情報漏えい対策が十分な他のファイルサーバ等において蓄積されたアクセスログによって示されるアクセス情報を基に学習されたものであってもよい。

【０１４２】

また、図２１は、本変形例による不審行動検知システムの動作例を示すフローチャートである。図２１に示す例は、図１３に示したアクセス行動予測ステップの動作例に比べて、最初の予測モデル読み出し動作（ステップＳ２０１）が、予測モデルの受信・読み出し動作（ステップＳ６０１）に変わっている点を除き、図１３に示したアクセス行動予測ステップの動作と同じである。すなわち、本変形例では、予測モデルを読み出しに際し、予測モデル受信部１１７で受信した予測モデルを読み出せばよい。

【０１４３】

例えば、ステップＳ６０１では、予測モデル受信部１１７は、ネットワーク経由で予測モデルを受信し、予測モデル記憶部１１０に書き込む。そして、予測スコア算出部１１１が、その予測モデルを予測モデル記憶部１１０から読み出す。

【０１４４】

本変形例によれば、自システムにおけるアクセスログの蓄積が十分でない場合やモデル生成に必要な処理能力が十分でない場合等であっても、精度のよい予測モデルを用いることができる。

【０１４５】

変形例４.

次に、本実施形態の第４変形例について説明する。これまで、学習・予測に用いる入力データとして、利用者データと、文書データの２つの入力データを想定して説明をしたが、アクセス行動学習ステップおよびアクセス行動予測ステップにおいて、３つ以上の入力データ（Ｎ入力データ）を処理させることも可能である。

【０１４６】

例えば、利用者データとして、次の３つのデータが存在するとする。すなわち、利用者データが、（ａ）いわゆる属性データ（図８に示した情報等の利用者自身に関するデータ）、（ｂ）ＳＮＳ等において生成されたデータであるＳＮＳデータ、（ｃ）当該利用者が所定のデータに対して行ったアクセス行動に関する統計値等の統計データ、に大別されるとする。

【０１４７】

このような場合に、システムは、上記３つのデータ各々から、上記のベクトル化と同じ方法で利用者特徴ベクトルを生成し、生成された３つの利用者特徴ベクトルをマージ（Ａ次元ベクトルとＢ次元ベクトルとＣ次元ベクトルと、・・・とをつなげてＡ＋Ｂ＋Ｃ＋・・・次元のベクトルに合成）して、１つの利用者特徴ベクトルとすればよい。文書データに関しても同様である。

【０１４８】

これにより、Ｎ個の入力データであっても、利用者とデータのどちらにより由来するかによって利用者データか文書データに分類し、マージを行うことで、２入力データに落とし込むことができる。

【０１４９】

変形例５.

次に、本実施形態の第５変形例について説明する。本実施形態のこれまでの説明では、アクセスログの特に指定した期間（予測期間）から抽出されるアクセス行動における＜利用者ＩＤ、文書ＩＤ＞の組に対して、不審行動か否かを判定した。しかし、予想対象とするアクセス行動は、このようなアクセスログによって示されるものに限定されない。例えば、実際に行われたアクセス行動に対してだけでなく、事前に、危険文書や危険利用者を予測することも可能である。ここで、危険文書は、ある特定の利用者または利用者群にと

って不審行動の対象となりやすい文書または文書群、より具体的には当該特定の利用者または利用者群がアクセスする可能性の低い文書または文書群をいう。また、危険利用者は、ある特定のデータまたはデータ群にとって不審行動の主体となりやすい利用者または利用者群、より具体的には当該特定のデータまたはデータ群にアクセスする可能性の低い利用者または利用者群をいう。危険文書や危険利用者を予め予測することにより、例えば、危険文書への特定利用者によるアクセスや、危険利用者による特定文書へのアクセスをあらかじめ制限する等の事前予防を実施できる。

【０１５０】

本変形例における危険利用者の予測方法は、例えば、アクセス行動予測ステップのステップＳ２０２で、アクセス行動予測対象リストを生成する際に、検査対象の利用者の利用者ＩＤ（特定利用者ＩＤ）に対して全ての文書ＩＤを組み合わせたものを、アクセス行動予測対象リストに含ませればよい。なお、予測に用いる入力データとして、利用者ＩＤおよび文書ＩＤから得られる情報以外の情報（例えば、アクセス時間帯等）を含む場合には、特定利用者ＩＤに対して利用者データ以外の入力データが取り得る全ての値のパターンを組み合わせたものを、アクセス行動予測対象リストに含ませればよい。

【０１５１】

そして、そのようにして生成したアクセス行動予測対象リストを用いて、ステップＳ２０３以降を実行すればよい。その結果、１つでも不審行動と判定された組があれば、その組に含まれる特定利用者ＩＤが示す利用者を、少なくともその組が示すアクセス行動における危険利用者とみなしてもよい。

【０１５２】

同様に、本変形例における危険文書を予測するには、例えば、アクセス行動予測ステップのステップＳ２０２で、アクセス行動予測対象リストを生成する際に、検査対象の文書の文書ＩＤ（特定文書ＩＤ）に対して全ての利用者ＩＤを組み合わせたものを、アクセス行動予測対象リストに含ませればよい。なお、予測に用いる入力データとして、利用者ＩＤおよび文書ＩＤから得られる情報以外の情報（例えば、アクセス時間帯等）を含む場合、該特定文書ＩＤに対して文書データ以外の入力データが取り得る全ての値のパターンを組み合わせたものを、アクセス行動予測対象リストに含ませればよい。

【０１５３】

そして、そのようにして生成したアクセス行動予測対象リストを用いて、ステップＳ２０３以降を実行すればよい。その結果、１つでも不審行動と判定された組があれば、その組に含まれる特定文書ＩＤが示す文書を、少なくともその組が示すアクセス行動における危険文書とみなしてもよい。

【０１５４】

また、システムは、危険利用者や危険文書が検出された場合、不審行動通知ステップの動作を実行してもよい。

【０１５５】

以上、本実施形態および実施例を参照して本願発明を説明したが、本願発明は上記実施形態および実施例に限定されるものではない。本願発明の構成や詳細には、本願発明のスコープ内で当業者が理解し得る様々な変更をすることができる。

【０１５６】

例えば、本願発明の特徴の１つは、データアクセスに関する過去の利用者の行動を示すデータをもとに、機械学習を行い、未知のデータアクセス行動に対して不審行動か否かを判定する点にある。上記の説明の多くでは、２入力（アクセスログから得られる、利用者データと文書データの１対１の組合せ）に対して成否のラベルを付けて学習を行う例を示している。しかし、本願発明の目的の１つとして、機械学習による行動ベースのアクセス制御ができればよいので、学習に用いる入力は上記に限られない。また、監視対象も、一企業等の情報システム部門で管理されるようなファイルサーバに限られない。

【０１５７】

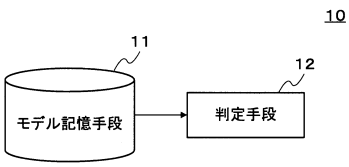
入力データに含まれるとして好ましい項目の一例として、データアクセス行動に関する

下記の５Ｗ１Ｈに相当する情報が挙げられる。
【０１５８】
ＷＨＯ：利用者のプロフィール（氏名、年齢、役職、職務、健康状態、上司評価、など）
ＷＨＥＮ：利用者がデータにアクセスした日時（平日、休日、日中、夜間、など）
ＷＨＥＲＥ：利用者がデータにアクセスした場所（ファイルサーバ、データベース、ＳＮＳ、など）
ＷＨＡＴ：利用者がアクセスしたデータ（タイトル、プロパティ、内容、など）
ＷＨＹ：利用者がデータにアクセスした理由（読込、書込、コピー、削除、など）
ＨＯＷ：利用者がデータにアクセスした方法（アクセス端末、アクセス経路、など）
【０１５９】
　また、例えば、第２の実施形態において利用者データ前処理部１０３や文書データ前処理部１０４が生成するベクトルの次元数がそれほど大きくない場合は、特徴抽出部（利用者属性特徴抽出部１０７、文書属性特徴抽出部１０８）を省略してもよい。
【０１６０】
　また、上記の各実施形態は以下の付記のようにも記載できる。
【０１６１】
　（付記１）データに対する利用者の行動であるデータアクセス行動に関するアクセス情報であって、データにアクセスする利用者に由来する第１の情報と、アクセスされるデータに由来する第２の情報とを含むアクセス情報と、不審行動または正常行動との関係を示すアクセス行動モデルを記憶するモデル記憶手段と、アクセス行動モデルに基づいて、任意のデータアクセス行動が不審行動であるか否かを判定する判定手段とを備えたことを特徴とする情報処理装置。
【０１６２】
　（付記２）アクセス情報は、第１の情報として、アクセスする利用者、アクセスされる時間、アクセス種別もしくはアクセス方法に関する情報を含む、または、第２の情報として、データ自体もしくはデータの格納場所に関する情報を含む付記１に記載の情報処理装置。
【０１６３】
　（付記３）アクセス情報は、アクセスする利用者に関する情報として、当該利用者が生成したテキストに関する情報もしくは当該利用者が所定のデータに対して行ったアクセス行動に関する統計値を含む、または、データ自体に関する情報として、当該データの内容に関する情報もしくは当該データに対して行われたアクセス行動に関する統計値を含む付記２に記載の情報処理装置。
【０１６４】
　（付記４）アクセス情報と、アクセス情報が示すデータアクセス行動が不審行動であるか否かを示す情報とを学習データに用いて、機械学習によりアクセス行動モデルを生成する学習手段を備えた付記１から付記３のうちのいずれかに記載の情報処理装置。
【０１６５】
　（付記５）ファイルサーバによって管理されているファイルを、対象データとする情報処理装置であって、モデル記憶手段は、所定のファイルに対するアクセス履歴に含まれるアクセス行動のうち指定された期間におけるアクセス行動に関するアクセス情報と、アクセス行動が不審行動か否かを判別可能な情報とを用いて機械学習されたアクセス行動モデルを記憶する付記１から付記４のうちのいずれかに記載の情報処理装置。
【０１６６】
　（付記６）アクセス情報から、各々が多次元の数値からなる２以上の数値ベクトルを生成する数値ベクトル生成手段を備え、モデル記憶手段は、２以上の数値ベクトルの組と、不審行動または正常行動との関係を示すアクセス行動モデルとの関係を示すアクセス行動モデルを記憶し、判定手段は、アクセス行動モデルを用いて算出される、指定されたアクセス情報から生成される２以上の数値ベクトルの組に対する不審行動または正常行動の確度に基づいて、アクセス情報によって示されるデータアクセス行動が不審行動であるか否

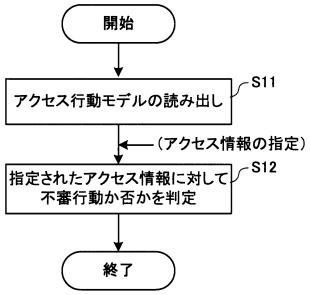
かを判定する付記１から付記５のうちのいずれかに記載の情報処理装置。
【０１６７】
（付記７）数値ベクトル生成手段として、アクセス情報に含まれる第１情報から、多次元の数値からなる第１数値ベクトルを生成する第１数値ベクトル生成手段と、アクセス情報に含まれる第２情報から、多次元の数値からなる第２数値ベクトルを生成する第２数値ベクトル生成手段とを備え、モデル記憶手段は、第１数値ベクトルと第２数値ベクトルとの組と、不審行動または正常行動との関係を示すアクセス行動モデルを記憶し、判定手段は、アクセス行動モデルを用いて算出される、指定されたアクセス情報に含まれる第１情報および第２情報から生成される第１数値ベクトルと第２数値ベクトルの組に対する不審行動または正常行動の確度に基づいて、アクセス情報によって示されるデータアクセス行動が不審行動であるか否かを判定する付記６に記載の情報処理装置。
【０１６８】
（付記８）アクセス行動モデルに基づいて、不審行動に該当するアクセス行動が行われる危険性があるデータを予測する危険データ予測手段を備えた付記１から付記７のうちのいずれかに記載の情報処理装置。
【０１６９】
（付記９）アクセス行動モデルに基づいて、不審行動に該当するデータアクセス行動を行う危険性がある利用者を予測する危険利用者予測手段を備えた付記１から付記８のうちのいずれかに記載の情報処理装置。
【０１７０】
（付記１０）判定手段による判定結果に基づいて、アクセス権限を変更するアクセス権限変更手段を備えた付記１から付記９のうちのいずれかに記載の情報処理装置。
【０１７１】
（付記１１）判定手段による判定結果を基に、実際に行われたデータアクセス行動から不審行動を検知する不審行動検知手段と、不審行動が検知されると、管理者に通知を行う通知手段とを備えた付記１から付記１０のうちのいずれかに記載の情報処理装置。
【０１７２】
（付記１２）データに対する利用者の行動であるデータアクセス行動に関するアクセス情報であって、データにアクセスする利用者に由来する第１の情報と、アクセスされるデータに由来する第２の情報とを含むアクセス情報と、アクセス情報が示すデータアクセス行動が不審行動か否かを判別可能な情報とを学習データに用いて、機械学習により、任意のアクセス情報と不審行動または正常行動との関係を示すアクセス行動モデルを生成する学習手段と、アクセス行動モデルを記憶するモデル記憶手段と、アクセス行動モデルに基づいて、任意のデータアクセス行動が不審行動であるかか否かを判定する判定手段と、判定結果を基に、実際に行われたデータアクセス行動から不審行動を検知する不審行動検知手段とを備えたことを特徴とする不審行動検知システム。
【０１７３】
（付記１３）情報処理装置が、データに対する利用者の行動であるデータアクセス行動に関するアクセス情報であって、データにアクセスする利用者に由来する第１の情報と、アクセスされるデータに由来する第２の情報とを含むアクセス情報と、不審行動または正常行動との関係を示すアクセス行動モデルに基づいて、任意のデータアクセス行動が不審行動であるか否かを判定することを特徴とする不審行動検知方法。
【０１７４】
（付記１４）コンピュータに、データに対する利用者の行動であるデータアクセス行動に関するアクセス情報であって、データにアクセスする利用者に由来する第１の情報と、アクセスされるデータに由来する第２の情報とを含むアクセス情報と、不審行動または正常行動との関係を示すアクセス行動モデルに基づいて、任意のデータアクセス行動が不審行動であるか否かを判定させる処理を実行させるための不審行動検知プログラム。
【０１７５】
この出願は、２０１５年１０月１３日に出願された日本特許出願２０１５－２０２２８

０を基礎とする優先権を主張し、その開示の全てをここに取り込む。
【産業上の利用可能性】
【０１７６】
本発明は、入力データから利用者およびデータに関する特徴量を抽出してモデル学習を行う特徴から、例えば、不審行動の検知に高い精度を有する予測モデルのみを提供するといったビジネスモデルも考えられる。
【符号の説明】
【０１７７】
１０、１００　不審行動検知システム
１１　モデル記憶手段
１２　判定手段
１３　学習手段
１４　不審行動検知手段
１５　通知手段
１６　数値ベクトル生成手段
１６１　第１数値ベクトル生成手段
１６２　第２数値ベクトル生成手段
１７　危険利用者予測手段
１８　危険データ予測手段
１９　アクセス権限変更手段
１０１　利用者データ記憶部
１０２　文書データ記憶部
１０３　利用者データ前処理部
１０４　文書データ前処理部
１０５　アクセスログ記憶部
１０６　アクセスログ前処理部
１０７　利用者属性特徴抽出部
１０８　文書属性特徴抽出部
１０９　アクセス実績学習部
１１０　予測モデル記憶部
１１１　予測スコア算出部
１１２　予測スコア記憶部
１１３　不審行動通知部
１１４　アクセス権限制御部
１１５　アクセス権限記憶部
１１６　アクセス権限制御画面部
１１７　予測モデル受信部

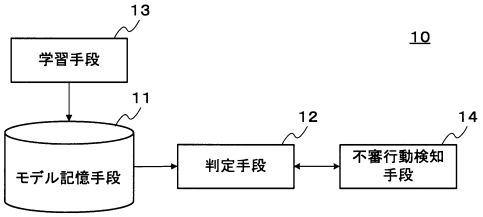
【図１】



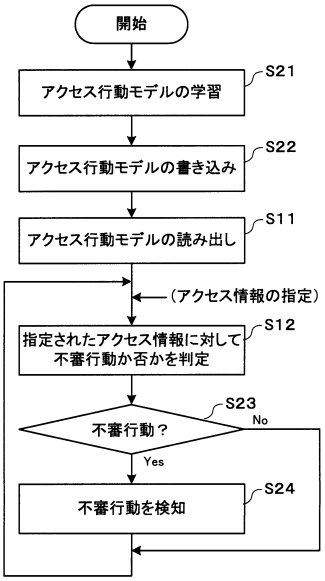
【図２】



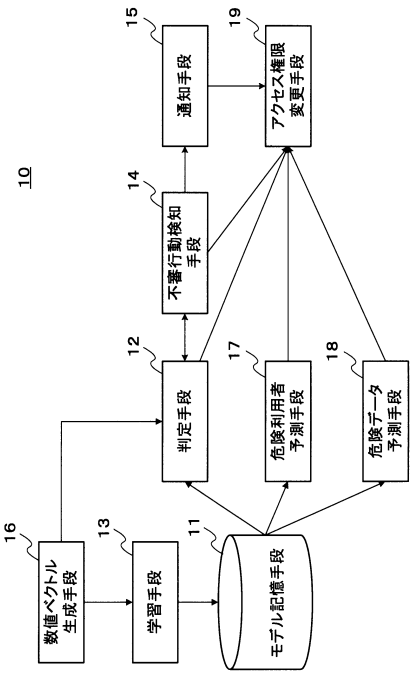
【図３】



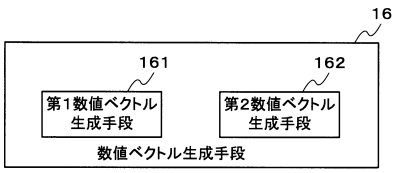
【図４】



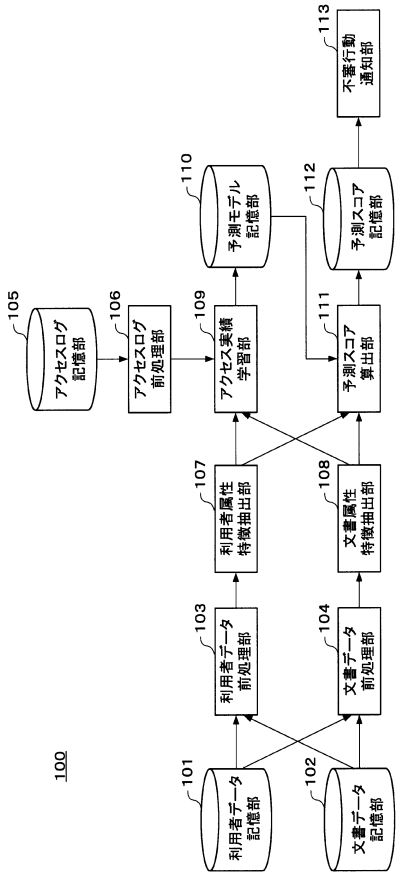
【図５】



【図６】



【図７】



【図８】

利用者ID	氏名	年齢	性別	役職	担当業務	業績評価	キャリア面接結果
U01	山田 一郎	22	男	担当	開発	A	現在の担当業務にやりがいを感じている。
U02	山下 花子	28	女	担当	営業	S	強い情熱を持って、担当業務に取り組んでいる。
U03	山本 二郎	48	男	部長	経理・財務	C	長年、担当業務一筋でマンネリ化。 専門スキルを活かしてキャリアチェンジを志向。
U04	松田 三郎	36	男	課長	法務	A	30歳の節目に現職に転職して以来、担当業務一筋。 日々楽しく業務している。
U05	松下 紀子	30	女	係長	知財	B	30歳の節目に係長に昇格。 仕事の裁量も増えて毎日忙しく仕事をこなしている。

【図９】

文書ID	文書種別	アクセス権	作成日時	更新日時	文書内容
D01	営業資料	営業職かつ担当以上は閲覧可能	2015/04/01	2015/05/01	顧客であるA社向けの営業リーフレットです。…
D02	技術資料	技術職かつ担当以上は閲覧可能	2013/04/01	2015/04/01	2015年6月リリースのB製品の基本設計書です。…
D03	決算資料	財務職かつ担当以上は閲覧可能	2000/01/01	2015/03/30	2014年度の決算報告書です。企業秘密です。…
D04	顧客資料	部長以上は閲覧可能	2000/01/01	2015/05/18	顧客データを含みます。企業秘密、取扱注意です。…
D05	広報資料	担当以上は閲覧可能	2014/02/01	2014/04/01	2014年4月に社外公表した広報資料です。…

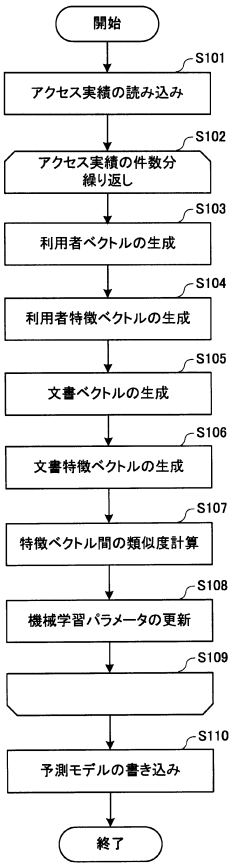
【図１０】

アクセス利用者ID	アクセス文書ID	アクセス日時
U01	D02	2015/05/16
U02	D01	2015/05/16
U03	D03	2015/05/16
U04	D05	2015/05/16
U01	D02	2015/05/17
U02	D01	2015/05/17
U03	D03	2015/05/17
U04	D05	2015/05/17
U01	D02	2015/05/18
U02	D01	2015/05/18
U03	D04	2015/05/18
U04	D05	2015/05/18

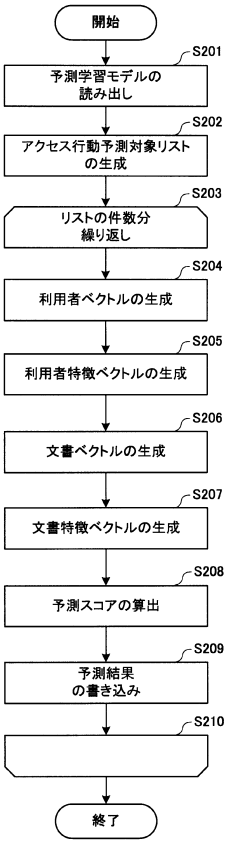
【図１１】

アクセス者ID	アクセス文書ID	予測スコア
U01	D02	0.9
U02	D01	0.8
U03	D04	0.1
U04	D05	0.7

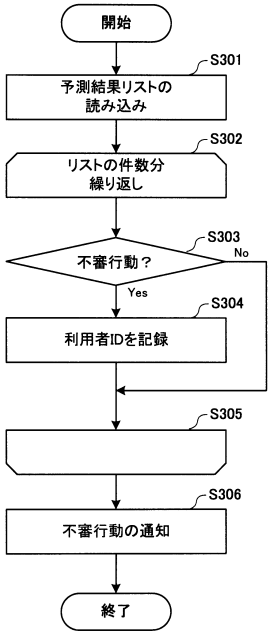
【図１２】



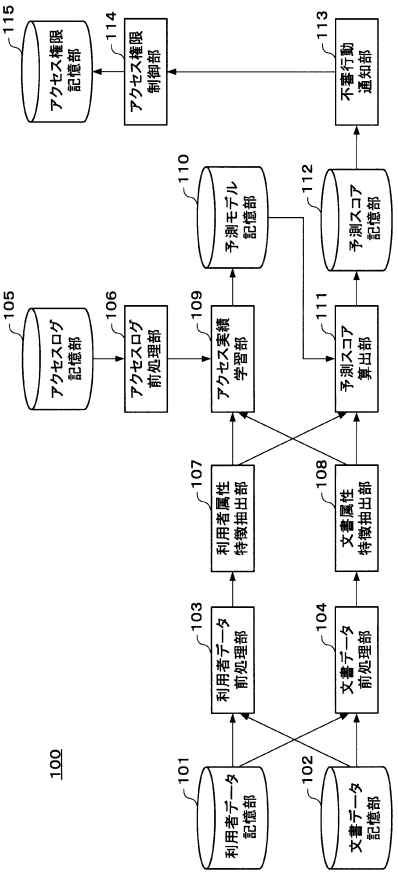
【図１３】



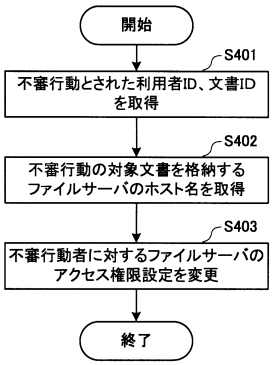
【図１４】



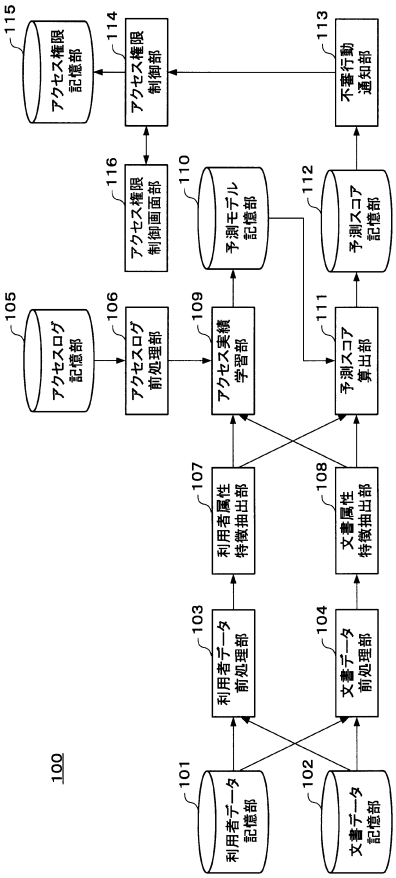
【図１５】



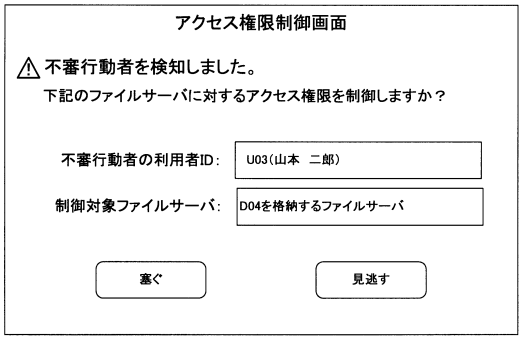
【図１６】



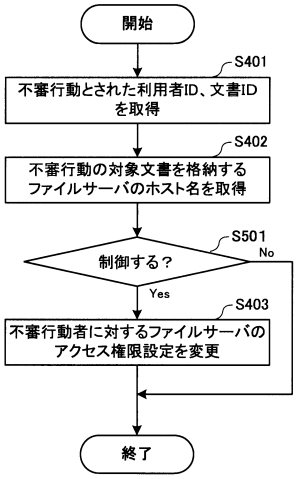
【図１７】



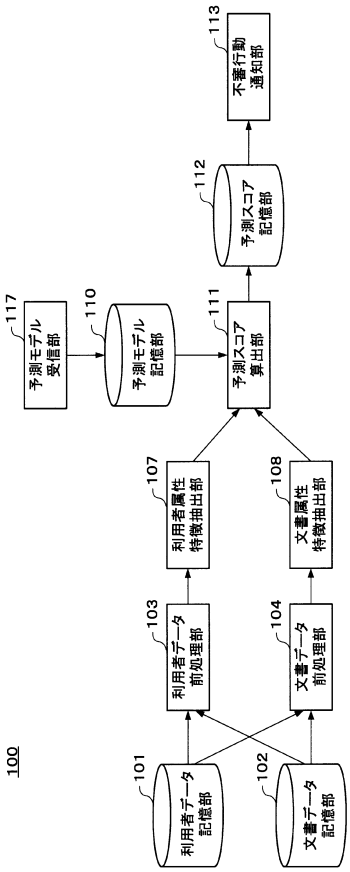
【図１８】



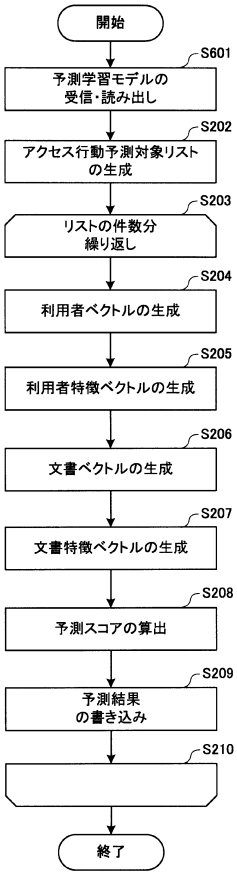
【図１９】



【図２０】


100
101 利用者データ記憶部
102 文書データ記憶部
103 利用者データ前処理部
104 文書データ前処理部
107 利用者属性特徴抽出部
108 文書属性特徴抽出部
110 予測モデル記憶部
111 予測スコア算出部
112 予測スコア記憶部
113 不審行動通知部
117 予測モデル受信部

【図２１】


開始
S601
予測学習モデルの受信・読み出し
S202
アクセス行動予測対象リストの生成
S203
リストの件数分繰り返し
S204
利用者ベクトルの生成
S205
利用者特徴ベクトルの生成
S206
文書ベクトルの生成
S207
文書特徴ベクトルの生成
S208
予測スコアの算出
S209
予測結果の書き込み
S210
終了

フロントページの続き

(56)参考文献 特開２００８－１５８９５９（ＪＰ，Ａ）
特開２０１０－００９２３９（ＪＰ，Ａ）
特開２００４－３１２０８３（ＪＰ，Ａ）
特開２００５－１９００６６（ＪＰ，Ａ）
特開２０００－１４８２７６（ＪＰ，Ａ）
特開２０１１－１３８２９８（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　２１／５５
Ｇ０６Ｆ　２１／６２
Ｇ０６Ｎ　２０／００