(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)公表番号
特表2024-513569
(P2024-513569A)
(43)公表日　令和6年3月26日(2024.3.26)

(51)国際特許分類　ＦＩ
Ｇ０６Ｆ　40/44　(2020.01)　Ｇ０６Ｆ　40/44

審査請求　未請求　予備審査請求　未請求　（全29頁）

(21)出願番号　特願2023-561710(P2023-561710)
(86)(22)出願日　令和4年3月28日(2022.3.28)
(85)翻訳文提出日　令和5年11月30日(2023.11.30)
(86)国際出願番号　PCT/EP2022/058130
(87)国際公開番号　WO2022/214348
(87)国際公開日　令和4年10月13日(2022.10.13)
(31)優先権主張番号　17/301,641
(32)優先日　令和3年4月9日(2021.4.9)
(33)優先権主張国・地域又は機関
米国(US)
(81)指定国・地域　AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,
最終頁に続く

(71)出願人　312016539
ビットディフェンダー アイピーアール マネジメント リミテッド
キプロス共和国、ニコシア、２０１２、アクロポレオス、５９－６１、サード・フロア、フラット／オフィス ３０２
(74)代理人　100118902
弁理士　山本 修
(74)代理人　100106208
弁理士　宮前 徹
(74)代理人　100196508
弁理士　松尾 淳一
(72)発明者　アンドレイ・エム，マノラチェ
ルーマニア国 ６１５２００ トゥルグ・ネアムツ，ジュデトゥル・ネアムツ，ス
最終頁に続く

(54)【発明の名称】　異常検出システムおよび方法

(57)【要約】
いくつかの実施形態が、自然言語処理やコンピュータセキュリティなどの適用分野での異常検出のために人工知能システム（たとえば、ディープニューラルネットワークのセット）をトレーニングする新規な手順を利用する。トレーニングコーパスから選択されたトークンシーケンスは、シーケンスアナライザに供給される前に、複数の所定のシーケンス変換のうちの少なくとも１つに従ってひずめられる。さらに、シーケンスアナライザは、それぞれの入力トークンシーケンスを生成するためにどの変換が使用されたかを正しく推測するようにトレーニングされる。

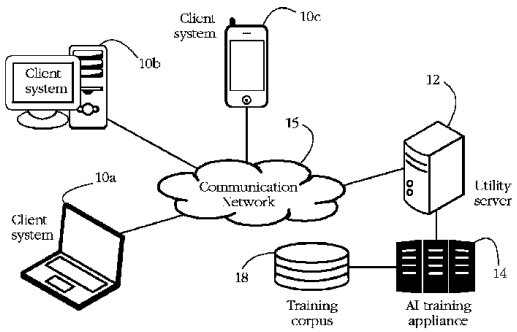

FIG. 1

【特許請求の範囲】
【請求項１】
コンピュータで実装された異常検出方法であって、前記方法は、コンピュータシステムの少なくとも１つのハードウェアプロセッサを利用して、
トークンシーケンスのトレーニングコーパスからトレーニング・トークン・シーケンスを選択することに応答して、および、所定の複数のシーケンス変換から変換を選択することに応答して、選択された前記変換を前記トレーニング・トークン・シーケンスに適用して、修正後トークンシーケンスを生成するステップと、
調節可能パラメータのセットを有し、前記修正後トークンシーケンスに従って変換予測標識を決定するように構成されたシーケンスアナライザを実行するステップであって、前記変換予測標識は、選択された前記変換が適用されて前記修正後トークンシーケンスが生成された可能性を示す、シーケンスアナライザを実行するステップと、
前記予測標識を決定することに応答して、前記変換予測標識に従って調節可能パラメータの前記セットのうちの少なくとも１つのパラメータを調節するステップと、
前記少なくとも１つのパラメータを調節することに応答して、前記シーケンスアナライザを利用してターゲット・トークン・シーケンスが異常であるかどうかを判定するステップと
を実施することを含む、コンピュータで実装された異常検出方法。
【請求項２】
請求項１に記載の方法であって、選択された前記変換を適用するステップは、前記トレーニング・トークン・シーケンスの選択されたトークンを代替トークンと置き換えるステップを含む、方法。
【請求項３】
請求項２に記載の方法であって、
調節可能パラメータの別のセットを有し、前記トレーニング・トークン・シーケンスに従って前記代替トークンを生成するように構成されたトークンジェネレータを実行するステップと、
前記変換予測標識を決定することに応答して、前記変換予測標識に従って調節可能パラメータの前記別のセットの別のパラメータを調節するステップと
をさらに含む方法。
【請求項４】
請求項１に記載の方法であって、
選択された前記変換を適用するステップは、前記トレーニング・トークン・シーケンスの選択されたトークンを削除するステップと、前記トレーニング・トークン・シーケンス内に追加のトークンを挿入するステップと、前記トレーニング・トークン・シーケンスのトークンの選択されたサブセットを並べ替えるステップとを含むグループから選択された項目を含む、方法。
【請求項５】
請求項１に記載の方法であって、
前記シーケンスアナライザは、前記修正後トークンシーケンスに従ってトークン予測標識を決定するようにさらに構成され、前記トークン予測標識は、前記修正後トークンシーケンスの選択されたトークンが、選択された前記変換の前記適用によって変更された可能性を示し、
前記少なくとも１つの調節可能パラメータを調節するステップは、前記トークン予測標識にさらに従って前記少なくとも１つの調節可能パラメータを調節するステップを含む、方法。
【請求項６】
請求項１に記載の方法であって、
前記トレーニング・トークン・シーケンスおよびターゲット・トークン・シーケンスは、自然言語で構築されたテキストを含み、

前記方法は、前記ターゲット・トークン・シーケンスが異常であるかどうかを判定することに応答して、前記ターゲット・トークン・シーケンスが異常であるとき、前記ターゲット・トークン・シーケンスの作成者が前記トレーニング・トークン・シーケンスの作成者とは異なると判定するステップをさらに含む、
方法。
【請求項７】
請求項１に記載の方法であって、
前記トレーニング・トークン・シーケンスおよびターゲット・トークン・シーケンスは、自然言語で構築されたテキストを含み、
前記方法は、前記ターゲット・トークン・シーケンスが異常であるかどうかを判定することに応答して、前記ターゲット・トークン・シーケンスが異常であるとき、前記ターゲット・トークン・シーケンスの主題が前記トレーニング・トークン・シーケンスの主題とは異なると判定するステップをさらに含む、
方法。
【請求項８】
請求項１に記載の方法であって、
前記トレーニング・トークン・シーケンスおよびターゲット・トークン・シーケンスは、自然言語で構築されたテキストを含み、
前記方法は、前記ターゲット・トークン・シーケンスが異常であるかどうかを判定することに応答して、前記ターゲット・トークン・シーケンスが異常であるとき、前記ターゲット・トークン・シーケンスがマシンで生成されたと判定するステップをさらに含む、
方法。
【請求項９】
請求項１に記載の方法であって、
前記トレーニングコーパスは、選択基準に従って選択されたテキストフラグメントを含み、
前記方法は、前記ターゲット・トークン・シーケンスが異常であるかどうかを判定することに応答して、前記ターゲット・トークン・シーケンスが異常であるとき、前記ターゲット・トークン・シーケンスが前記選択基準を満たさないと判定するステップをさらに含む、
方法。
【請求項１０】
請求項１に記載の方法であって、
前記トレーニング・トークン・シーケンスおよびターゲット・トークン・シーケンスは、コンピューティングイベントのシーケンスを含み、
前記方法は、前記ターゲット・トークン・シーケンスが異常であるかどうかを判定することに応答して、前記ターゲット・トークン・シーケンスが異常であるとき、前記ターゲット・トークン・シーケンスがコンピュータセキュリティ脅威を示すと判定するステップをさらに含む、
方法。
【請求項１１】
少なくとも１つのハードウェアプロセッサを備えるコンピュータシステムであって、前記少なくとも１つのハードウェアプロセッサは、
トークンシーケンスのトレーニングコーパスからトレーニング・トークン・シーケンスを選択することに応答して、および、所定の複数のシーケンス変換から変換を選択することに応答して、選択された前記変換を前記トレーニング・トークン・シーケンスに適用して、修正後トークンシーケンスを生成することと、
調節可能パラメータのセットを有し、前記修正後トークンシーケンスに従って変換予測標識を決定するように構成されたシーケンスアナライザを実行することであって、前記変換予測標識は、選択された前記変換が適用されて前記修正後トークンシーケンスが生成さ

れた可能性を示す、シーケンスアナライザを実行することと、
前記予測標識を決定することに応答して、前記変換予測標識に従って調節可能パラメータの前記セットのうちの少なくとも１つのパラメータを調節することと、
前記少なくとも１つのパラメータを調節することに応答して、前記シーケンスアナライザを実行してターゲット・トークン・シーケンスが異常であるかどうかを判定することとを行うように構成された、コンピュータシステム。
【請求項１２】
請求項１１に記載のコンピュータシステムであって、選択された前記変換を適用することは、前記トレーニング・トークン・シーケンスの選択されたトークンを代替トークンと置き換えることを含む、コンピュータシステム。
【請求項１３】
請求項１２に記載のコンピュータシステムであって、前記少なくとも１つのハードウェアプロセッサは、
調節可能パラメータの別のセットを有し、前記トレーニング・トークン・シーケンスに従って前記代替トークンを生成するように構成されたトークンジェネレータを実行することと、
前記変換予測標識を決定することに応答して、前記変換予測標識に従って調節可能パラメータの前記別のセットの別のパラメータを調節することと
を行うようにさらに構成される、コンピュータシステム。
【請求項１４】
請求項１１に記載のコンピュータシステムであって、
選択された前記変換を適用することは、前記トレーニング・トークン・シーケンスの選択されたトークンを削除することと、前記トレーニング・トークン・シーケンス内に追加のトークンを挿入することと、前記トレーニング・トークン・シーケンスのトークンの選択されたサブセットを並べ替えることとを含むグループから選択された項目を含む、コンピュータシステム。
【請求項１５】
請求項１１に記載のコンピュータシステムであって、
前記シーケンスアナライザは、前記修正後トークンシーケンスに従ってトークン予測標識を決定するようにさらに構成され、前記トークン予測標識は、前記修正後トークンシーケンスの選択されたトークンが選択された前記変換の前記適用によって変更された可能性を示し、
前記少なくとも１つの調節可能パラメータを調節することは、前記トークン予測標識にさらに従って前記少なくとも１つの調節可能パラメータを調節することを含む、
コンピュータシステム。
【請求項１６】
請求項１１に記載のコンピュータシステムであって、
前記トレーニング・トークン・シーケンスおよびターゲット・トークン・シーケンスは、自然言語で構築されたテキストを含み、
前記方法は、前記ターゲット・トークン・シーケンスが異常であるかどうかを判定することに応答して、前記ターゲット・トークン・シーケンスが異常であるとき、前記ターゲット・トークン・シーケンスの作成者が前記トレーニング・トークン・シーケンスの作成者とは異なると判定することをさらに含む、
コンピュータシステム。
【請求項１７】
請求項１１に記載のコンピュータシステムであって、
前記トレーニング・トークン・シーケンスおよびターゲット・トークン・シーケンスは、自然言語で構築されたテキストを含み、
前記方法は、前記ターゲット・トークン・シーケンスが異常であるかどうかを判定することに応答して、前記ターゲット・トークン・シーケンスが異常であるとき、前記ターゲ

ット・トークン・シーケンスの主題が前記トレーニング・トークン・シーケンスの主題とは異なると判定することをさらに含む、
コンピュータシステム。
【請求項１８】
請求項１１に記載のコンピュータシステムであって、
前記トレーニング・トークン・シーケンスおよびターゲット・トークン・シーケンスは、自然言語で構築されたテキストを含み、
前記方法は、前記ターゲット・トークン・シーケンスが異常であるかどうかを判定することに応答して、前記ターゲット・トークン・シーケンスが異常であるとき、前記ターゲット・トークン・シーケンスがマシンで生成されたと判定することをさらに含む、
コンピュータシステム。
【請求項１９】
請求項１１に記載のコンピュータシステムであって、
前記トレーニングコーパスは、選択基準に従って選択されたテキストフラグメントを含み、
前記方法は、前記ターゲット・トークン・シーケンスが異常であるかどうかを判定することに応答して、前記ターゲット・トークン・シーケンスが異常であるとき、前記ターゲット・トークン・シーケンスが前記選択基準を満たさないと判定することをさらに含む、
コンピュータシステム。
【請求項２０】
請求項１１に記載のコンピュータシステムであって、
前記トレーニング・トークン・シーケンスおよびターゲット・トークン・シーケンスは、コンピューティングイベントのシーケンスを含み、
前記方法は、前記ターゲット・トークン・シーケンスが異常であるかどうかを判定することに応答して、前記ターゲット・トークン・シーケンスが異常であるとき、前記ターゲット・トークン・シーケンスがコンピュータセキュリティ脅威を示すと判定することをさらに含む、
コンピュータシステム。
【請求項２１】
命令を記憶する非一時的コンピュータ可読媒体であって、前記命令は、コンピュータシステムの少なくとも１つのハードウェアプロセッサによって実行されるとき、前記コンピュータシステムに、
トークンシーケンスのトレーニングコーパスからトレーニング・トークン・シーケンスを選択することに応答して、および、所定の複数のシーケンス変換から変換を選択することに応答して、選択された前記変換を前記トレーニング・トークン・シーケンスに適用して、修正後トークンシーケンスを生成することと、
調節可能パラメータのセットを有し、前記修正後トークンシーケンスに従って変換予測標識を決定するように構成されたシーケンスアナライザを実行することであって、前記変換予測標識は、選択された前記変換が適用されて前記修正後トークンシーケンスが生成された可能性を示す、シーケンスアナライザを実行することと、
前記予測標識を決定することに応答して、前記変換予測標識に従って調節可能パラメータの前記セットのうちの少なくとも１つのパラメータを調節することと、
前記少なくとも１つのパラメータを調節することに応答して、前記シーケンスアナライザを実行してターゲット・トークン・シーケンスが異常であるかどうかを判定することとを行わせる、非一時的コンピュータ可読媒体。
【発明の詳細な説明】
【技術分野】
【０００１】
[0001]本発明は人工知能に関し、詳細には自然言語処理およびコンピュータセキュリティの適用分野のための、データ内の異常を自動的に検出するためのシステムおよび方法

に関する。
【背景技術】
【０００２】
[0002]人工知能（ＡＩ）技術および機械学習技術は、とりわけパターン認識、自動分類、および異常検出などの適用分野に関して大量のデータを処理するためにますます使用されている。異常検出は、基準グループによって集合的に定義された標準または「正常」から大幅に逸脱する標本を識別することに相当する。異常検出は、複雑なデータのケースではかなりの技術的課題を提起し得、そのケースでは、正常性の意味および境界が事前に明確ではないことがあり、または事前に定義されないことがある。データから精巧なモデルを自動的に推論する能力を用いて、現代の人工知能システム（たとえばディープニューラルネットワーク）は、そのような課題に対して良好に動作することが示されている。
【０００３】
[0003]しかしながら、異常検出器をトレーニングするために機械学習を実装することは、それ自体の技術的課題のセットを提起する。従来の手法のいくつかでは、トレーニングが極度の計算コストを招くことがあり、非常に大規模なトレーニングコーパスを必要とすることがあり、不安定および／または非効率であることがある。したがって、自然言語処理およびコンピュータセキュリティの適用分野について異常検出器をトレーニングする新規な検出器アーキテクチャおよび新規な方法を開発することにかなりの関心が向けられている。
【発明の概要】
【０００４】
[0004]一態様によれば、コンピュータで実装された異常検出方法が、トークンシーケンスのトレーニングコーパスからトレーニング・トークン・シーケンスを選択することに応答して、かつ所定の複数のシーケンス変換から変換を選択することに応答して、コンピュータシステムの少なくとも１つのハードウェアプロセッサを利用して、選択された変換をトレーニング・トークン・シーケンスに適用して、修正後トークンシーケンスを生成するステップを含む。方法は、調節可能パラメータのセットを有し、修正後トークンシーケンスに従って変換予測標識を決定するように構成されたシーケンスアナライザを実行するステップであって、変換予測標識は、選択された変換が適用されて修正後トークンシーケンスが生成された可能性を示す、シーケンスアナライザを実行するステップをさらに含む。方法は、予測標識を決定することに応答して、変換予測標識に従って調節可能パラメータのセットのうちの少なくとも１つのパラメータを調節するステップと、少なくとも１つのパラメータを調節することに応答して、シーケンスアナライザを利用してターゲット・トークン・シーケンスが異常であるかどうかを判定するステップとをさらに含む。
【０００５】
[0005]別の態様によれば、コンピュータシステムが、トークンシーケンスのトレーニングコーパスからトレーニング・トークン・シーケンスを選択することに応答して、かつ所定の複数のシーケンス変換から変換を選択することに応答して、選択された変換をトレーニング・トークン・シーケンスに適用して、修正後トークンシーケンスを生成することを行うように構成された少なくとも１つのハードウェアプロセッサを備える。少なくとも１つのハードウェアプロセッサは、調節可能パラメータのセットを有し、修正後トークンシーケンスに従って変換予測標識を決定するように構成されたシーケンスアナライザを実行することであって、変換予測標識は、選択された変換が適用されて修正後トークンシーケンスが生成された可能性を示す、シーケンスアナライザを実行することを行うようにさらに構成される。少なくとも１つのハードウェアプロセッサは、予測標識を決定することに応答して、変換予測標識に従って調節可能パラメータのセットのうちの少なくとも１つのパラメータを調節することと、少なくとも１つのパラメータを調節することに応答して、シーケンスアナライザを利用してターゲット・トークン・シーケンスが異常であるかどうかを判定することを行うようにさらに構成される。
【０００６】

[0006]別の態様によれば、非一時的コンピュータ可読媒体が、コンピュータシステムの少なくとも１つのハードウェアプロセッサによって実行されるとき、コンピュータシステムに、トークンシーケンスのトレーニングコーパスからトレーニング・トークン・シーケンスを選択することに応答して、かつ所定の複数のシーケンス変換から変換を選択することに応答して、選択された変換をトレーニング・トークン・シーケンスに適用して、修正後トークンシーケンスを生成させることを行わせる命令を記憶する。命令はさらに、コンピュータシステムに、調節可能パラメータのセットを有し、修正後トークンシーケンスに従って変換予測標識を決定するように構成されたシーケンスアナライザを実行することであって、変換予測標識は、選択された変換が適用されて修正後トークンシーケンスが生成された可能性を示す、シーケンスアナライザを実行することを行わせる。命令はさらに、コンピュータシステムに、予測標識を決定することに応答して、変換予測標識に従って調節可能パラメータのセットのうちの少なくとも１つのパラメータを調節させることと、少なくとも１つのパラメータを調節することに応答して、シーケンスアナライザを利用してターゲット・トークン・シーケンスが異常であるかどうかを判定することとを行わせる。

【０００７】

[0007]以下の詳細な説明を読み、図面を参照するとき、本発明の前述の態様および利点をより良く理解されよう。

【図面の簡単な説明】

【０００８】

【図１】[0008]本発明のいくつかの実施形態による、異常を検出する際にユーティリティサーバと協働するクライアントシステムのセットを示す図である。

【図２】[0009]本発明のいくつかの実施形態による、異常検出器の例示的動作を示す図である。

【図３】[0010]本発明のいくつかの実施形態による、異常検出器の例示的トレーニングを示す図である。

【図４】[0011]本発明のいくつかの実施形態による、入力修正器の例示的動作を示す図である。

【図５】[0012]本発明のいくつかの実施形態による例示的トークン埋込み空間を示す図である。

【図６】[0013]いくつかの実施形態による例示的シーケンス変換を示し、図示される変換が、選択されたトークンの代表的ベクトルをナッジすることを含む図である。

【図７】[0014]本発明のいくつかの実施形態によるシーケンス分類器の例示的構造を示す図である。

【図８】[0015]本発明のいくつかの実施形態による、異常検出器のトレーニング中に実施されるステップの例示的シーケンスを示す図である。

【図９】[0016]本発明のいくつかの実施形態による、トレーニング済み異常検出器によって実施されるステップの例示的シーケンスを示す図である。

【図１０】[0017]本明細書で説明される方法のうちのいくつかを実施するように構成された例示的コンピューティングアプライアンスを示す図である。

【発明を実施するための形態】

【０００９】

[0018]以下の説明では、構造間のすべての記載の接続が直接的に動作可能な接続または中間構造を通じた間接的に動作可能な接続であり得ることを理解されたい。要素のセットは１つまたは複数の要素を含む。要素のどんな記載も少なくとも１つの要素を指すと理解されたい。複数の要素は少なくとも２つの要素を含む。別段に指定されていない限り、「または（ＯＲ）」のどんな使用も非排他的論理和を指す。別段に必要とされない限り、記載のどんな方法ステップも、必ずしも特定の図示される順序で実施される必要はない。第２の要素から導出された第１の要素（たとえばデータ）は、第２の要素と等しい第１の要素、ならびに第２の要素と、任意選択で他のデータとを処理することによって生成され

た第１の要素を包含する。パラメータに従って決定または判断を行うことは、パラメータに従って、かつ任意選択で他のデータに従って決定または判断を行うことを包含する。別段に指定されていない限り、いくつかの量／データの標識は、量／データ自体、または量／データ自体とは異なる標識であり得る。コンピュータプログラムは、タスクを実施するプロセッサ命令のシーケンスである。本発明のいくつかの実施形態で説明されるコンピュータプログラムは、スタンドアロン・ソフトウェア・エンティティまたは他のコンピュータプログラムのサブエンティティ（たとえば、サブルーチン、ライブラリ）であり得る。コンピュータ可読媒体は、磁気、光、および半導体記憶媒体（たとえば、ハードドライブ、光ディスク、フラッシュメモリ、ＤＲＡＭ）などの非一時的媒体、ならびに導電性ケーブルや光ファイバリンクなどの通信リンクを包含する。いくつかの実施形態によれば、本発明は、とりわけ、本明細書に記載の方法を実施するようにプログラムされるハードウェア（たとえば、１つまたは複数のプロセッサ）、ならびに本明細書に記載の方法を実施するための命令を符号化するコンピュータ可読媒体を備えるコンピュータシステムを提供する。

【００１０】

[0019]以下の説明は、必ずしも限定としてではなく、例として本発明の実施形態を示す。

[0020]図１は、本発明のいくつかの実施形態による、データ内の異常を検出するためにユーティリティサーバ１２と協働し得るクライアントシステム１０ａ～ｃの例示的セットを示す。本明細書では、異常とは、項目の基準集合／コーパスによって集合的に表される標準または「正常」から大幅に逸脱する項目を表すと理解されたい。この説明は、異常なテキストフラグメントやコンピューティングイベントシーケンスなどの異常なトークンシーケンスを検出することに焦点を当てる。そのような実施形態では、例示的な異常検出は、ターゲットテキストの作成者が基準テキストとは異なると判定することを含む。別の例示的な異常検出は、コンピューティングイベントのシーケンスがそれぞれのコンピュータの正常な挙動から逸脱し、恐らくはセキュリティ違反または悪意のあるソフトウェアの存在を示すと判定することを含む。いくつかの例示的な異常検出使用事例シナリオが以下で説明される。

【００１１】

[0021]クライアントシステム１０ａ～ｃは一般に、プロセッサ、メモリ、および通信インターフェースを有する任意の電子アプライアンスを表す。例示的クライアントシステム１０ａ～ｃは、とりわけ、パーソナルコンピュータ、企業メインフレームコンピュータ、サーバ、ラップトップ、タブレットコンピュータ、モバイル遠隔通信デバイス（たとえば、スマートフォン）、メディアプレーヤ、ＴＶ、ゲームコンソール、ホームアプライアンス、およびウェアラブルデバイス（たとえば、スマートウォッチ）を含む。図示されるクライアントシステムは通信ネットワーク１５によって相互接続され、通信ネットワーク１５は、ローカルエリアネットワーク（ＬＡＮ）および／またはインターネットなどの広域ネットワーク（ＷＡＮ）を含み得る。サーバ１２は一般に、互いに物理的に近接することがあり、または近接しないことがある、通信可能に結合されたコンピュータシステムのセットを表す。

【００１２】

[0022]図２は、本発明のいくつかの実施形態による例示的異常検出器２０の動作を示す。異常検出器２０は、ソフトウェアとして、すなわちメモリ内にロードされ、パーソナルコンピュータやスマートフォンなどのコンピューティングアプライアンスのハードウェアプロセッサによって実行されるとき、それぞれのアプライアンスにそれぞれのタスクを実施させる命令を含むコンピュータプログラムのセットとして実施され得る。しかしながら、そのような実施形態が限定を意味するわけではないことを当業者は理解されよう。むしろ、検出器２０は、ソフトウェアとハードウェアの任意の組合せとして実装され得る。たとえば、検出器２０の一部またはすべての機能が、フィールドプログラマブルゲートアレイ（ＦＰＧＡ）や他の特定用途向け集積回路（ＡＳＩＣ）などのファームウェアおよび

／または専用ハードウェアで実装され得る。それぞれのハードウェアモジュールは、それぞれの機能向けに高度に最適化され、たとえば、特定のバージョンのディープニューラルネットワークアーキテクチャを直接的に実装し、したがって汎用プロセッサに関して達成可能なものよりもかなり速い処理速度を可能にし得る。さらに、以下で説明されるように異常検出器２０および／または検出器２０をトレーニングするように構成されたコンピュータシステムの別個の構成要素が、別個であるが通信可能に結合されたマシン上で、かつ／または同一のコンピュータシステムの別個のハードウェアプロセッサ上で実行され得ることを当業者は理解されよう。

【００１３】

[0023]異常検出器２０は、ターゲット・トークン・シーケンス２２のコンピュータ可読符号化を受け取り、それに応答して、それぞれのトークンシーケンス２２が異常であるかどうかを示す異常標識２６を出力するように構成され得る。例示的トークンシーケンスは、とりわけ、英語や中国語などの自然言語で構築されたテキストのフラグメントなどのトークンの順序付き配列を含む。一般性を失うことなく、以下の説明は、主に自然言語処理の例に焦点を当て、例示的トークンは、とりわけ、個々の語、語句、文、数、句読点（たとえば、？！；：／（），．．．）、特殊文字（たとえば、＄＃％）、省略語（ＵＳＡ、ＬＯＬ、ＩＭＨＯなど）、ソーシャルメディアハンドル（たとえば、＠ＰＯＴＵＳ）、ハッシュタグ、エモーティコンを含み得る。本明細書で説明されるシステムおよび方法は、とりわけ、コンピューティングイベントのシーケンスやサウンドシーケンス（たとえば、音楽、音声）などの他のタイプのトークンシーケンスを処理することに適合され得ることを当業者は理解されよう。

【００１４】

[0024]例示的異常標識２６は、それぞれのターゲット・トークン・シーケンスが異常である可能性を示す数値スコアを含む。スコアはブール（たとえば、ＹＥＳ／ＮＯ）であり得、または所定の境界の間（たとえば、０から１の間）で徐々に変動し得る。そのような一例では、大きい値は、それぞれのシーケンスが異常である可能性が高いことを示す。代替の異常標識２６は、シーケンス２２が属する可能性が高いトークンシーケンスのカテゴリを示す分類ラベル（たとえば、異常、正常、未知、容疑）を含み得る。

【００１５】

[0025]１つの例示的シナリオでは、異常検出器２０の別個のインスタンスが、各クライアントシステム１０ａ～ｃ上で実行され得、したがって各クライアントは、それ自体の異常検出活動をローカルに独立して実施し得る。代替実施形態では、異常検出器２０はユーティリティサーバ１２上で実行され得、したがってユーティリティサーバ１２は、複数のクライアントシステム１０ａ～ｃの代わりに集中型異常検出活動を実施し得る。そのような実施形態では、サーバ１２は、各クライアントシステム１０ａ～ｃからターゲット・トークン・シーケンス２２の符号化を受信し、それぞれの異常標識２６をそれぞれのクライアントに返し得る。そのような一例では、クライアント１０ａ～ｃは、ユーティリティサーバ１２によって公開されるウェブインターフェースを介して異常検出サービスにアクセスし得る。

【００１６】

[0026]図３は、判断モジュール４４に接続されたシーケンスアナライザ４２などの異常検出器の例示的構成要素を示す。いくつかの実施形態では、シーケンスアナライザ４２は、基準トークンシーケンスのコーパス１８に関してトレーニングされたディープニューラルネットワークなどの人工知能（ＡＩ）システムを備える。自然言語処理のシナリオでは、コーパス１８は、自然言語（たとえば、英語）で書かれたテキストフラグメントの集合を含み得る。コーパス１８のより具体的な例は、特定の作成者によるテキストの集合、電子メッセージ（たとえば、ショートメッセージサービス－ＳＭＳメッセージ、ｅメール、ソーシャルメディアポストなど）の集合、特定の話題または関心のエリア（たとえば、ビジネスニュース、スポーツ、中東など）に関するテキストの集合、および特定のスタイル（たとえば、フィクション、詩、科学記事、ニュースなど）で書かれたテキストの集合

からなり得る。個々のコーパス項目は、たとえばメタデータを使用して、タグ付けされ、ラベル付けされ、かつ／または注釈付けされ得る。例示的メタデータは、項目の選択されたクラス／カテゴリ（たとえば、特定のユーザによって送られたｅメールメッセージ、金融ニュースなど）に対するメンバシップの標識を含み得る。コーパス１８は、当技術分野で周知の任意のフォーマットで、たとえばリレーショナルデータベース、単純リスト、またはＸＭＬもしくはＪＳＯＮフォーマットで指定された構造化データとして編成され、記憶され得る。

【００１７】

[0027]コーパス１８の内容は、通信の基準パターンまたは「正常な」パターンを集合的に定義し、いくつかの実施形態では、異常検出器２０は、それぞれの基準パターンの内部モデルを構築し、それに応答して、ターゲット・テキスト・フラグメントが学習済みのパターンに適合するか否かを判定するようにトレーニングされ得る。ターゲット・トークン・シーケンス２２が（コーパス１８に従って）「正常な」テキストに対応する基準パターンに適合しないことが判明したとき、シーケンス２２は異常であると見なされ、異常標識２６を介してそのようなものとしてレポートされ得る。

【００１８】

[0028]いくつかの実施形態では、異常検出器２０のトレーニングが、図１のＡＩトレーニングアプライアンス１４として示される、別々の専用コンピュータシステムによって実施される。アプライアンス１４は、ユーティリティサーバ１２および／またはクライアントシステム１０ａ～ｃに通信可能に結合され得、計算コストのかかるトレーニング手順を容易にするための、グラフィックス処理装置（ＧＰＵ）ファームなどの専用ハードウェアを備え得る。「トレーニング」という用語は通常、当技術分野では機械学習手順を示すために使用され、それによって、人工知能システム（たとえば、ニューラルネットワーク）に様々なトレーニング入力が提示され、それぞれの入力が生成する出力に従って人工知能システムが徐々に調整される。各トレーニング入力／バッチについて、トレーニングは、それぞれの入力を処理してトレーニング出力を生成することと、それぞれのトレーニング出力および／または入力に従って問題特有のユーティリティ関数の値を求めることと、それぞれのユーティリティ値に従ってそれぞれのＡＩシステムのパラメータのセットを調節することとを含み得る。パラメータを調節することは、ユーティリティ関数を最大化する（いくつかのケースでは、最小化する）ことを目標とし得る。ニューラルネットワークをトレーニングする一例では、調節可能パラメータはシナプス重みのセットを含み得、一方、ユーティリティ関数は、予想される出力または所望の出力からのトレーニング出力の逸脱を定量化し得る。そのような一例では、トレーニングは、それぞれのトレーニング入力に対応する所望の出力にトレーニング出力を近づけるように、シナプス重み、および場合によっては他のネットワークパラメータを調節することを含み得る。既知のトレーニングのフレーバは、とりわけ、教師あり、教師なし、自己教師あり、および強化学習を含む。いくつかの実施形態では、典型的な検出器２０の調節可能パラメータの数は、数千から数百万まで様々であり得る。トレーニングが成功すると、最適化された検出器パラメータ値２４（図２）のセットを生成し得、最適化された検出器パラメータ値２４は、クライアントシステム１０ａ～ｃおよび／またはユーティリティサーバ１２上で実行中の異常検出器２０のローカルインスタンスをインスタンス化するために使用され得る。

【００１９】

[0029]検出器２０のトレーニングが、図３に概略的に示されている。ＡＩトレーニングアプライアンス１４の同一のハードウェアプロセッサまたは物理マシン上で実行するために、図示される構成要素のすべてが必要であるわけではないことを当業者は理解されよう。

【００２０】

[0030]本発明のいくつかの実施形態は、トレーニングコーパス１８内に含まれるサンプルのうちの少なくともいくつかを、異常検出器２０内に供給する前にひずませ、次いで検出器２０をトレーニングして、適用されたひずみのタイプを識別する。図３に示される

一例では、入力修正器４０が、トレーニングコーパス１８から選択されたトレーニング・トークン・シーケンス３２を受け取り、トレーニングシーケンス３２にシーケンス変換３０の所定のセットのうちの少なくとも１つを適用した結果を含む修正後トークンシーケンス３４を出力するように構成される。

【００２１】

[0031]例示的シーケンス変換３０は、とりわけ、シーケンス３２内のトークンの選択されたサブセットを代替トークンと置き換えることと、シーケンス３２からトークンの選択されたサブセットを削除することと、トークンのセットをシーケンス３２内に挿入することと、シーケンス３２内のトークンの選択されたサブセットを並べ替えることとを含む。それぞれの変換による修正の目標とされるトークンのサブセットは、トレーニングシーケンス内の各トークンの位置に従って選択され得る。目標とされる位置はバイナリマスクによって示され得、０が、不変のままにされるトークンの位置をマークし、１が、それぞれの変換によって影響を受ける位置をマークする。たとえば、マスク［０　０　１　０　１］によって定義される並べ替え変換は、トークンシーケンス「Ｔｈｅｙ　ｗｅｒｅ　ｐｒｅｐａｒｅｄ　ｔｏ　ｌｅａｖｅ」を修正後トークンシーケンス「Ｔｈｅｙ　ｗｅｒｅ　ｌｅａｖｅ　ｔｏ　ｐｒｅｐａｒｅｄ」に変換し得、第３のトークンが第５のトークンと交換された。

【００２２】

[0032]代替実施形態では、変換３０によって目標とされるトークンは、それぞれのトークンのタイプに従って選択され得る。たとえば、いくつかの変換は、特定の品詞（たとえば、名詞、動詞、形容詞）または特定の文法的役割（たとえば、文の主語）を有するトークンを目標とし得る。１つのそのような例示的変換は、動詞を代替動詞または動詞句で置き換え得る。それぞれの代替トークンまたはトークンシーケンスは、ターゲットトークン／シーケンスの同意語または反意語となるように選択され得る。シーケンス変換３０のより高度な例は、パラフレージング、すなわち意味を保持しながらトークンシーケンス全体を代替シーケンスと置き換えることを含み得る。パラフレージング変換の一例は、トレーニングシーケンス「Ｋｉｄｓ　ｂｙ　ｔｈｅ　ｌａｋｅ　ｗｅｒｅ　ｂｅｉｎｇ　ｅａｔｅｎ　ａｌｉｖｅ　ｂｙ　ｍｏｓｑｕｉｔｏｅｓ」を修正後シーケンス「Ｍｏｓｑｕｉｔｏｅｓ　ｆｅｒｏｃｉｏｕｓｌｙ　ａｔｔａｃｋｅｄ　ｔｈｅ　ｋｉｄｓ　ｂｙ　ｔｈｅ　ｌａｋｅ」と置き換えることを含む。

【００２３】

[0033]同様の変換が、シーケンス３２の各トークンが個々のコンピューティングイベントを含む、コンピュータセキュリティ実施形態で適用され得る。たとえば、例示的変換３０が、タイプ「ｃｒｅａｔｅ　ｐｒｏｃｅｓｓ」のトークンをトレーニングシーケンス３２から除去し得る。そのような実施形態でのパラフレージングの同等物が、イベントのターゲットシーケンスを、それぞれのコンピュータシステムを同一の最終状態にすることになるイベントの代替シーケンスと置き換えることを含み得る。

【００２４】

[0034]図４は、本発明のいくつかの実施形態による、複数の事前定義されたシーケンス変換３０を実装する入力修正器４０の例示的動作を示す。トレーニングシーケンス３２はトークン３５ａ～ｅを含み、現在の例では、トークン３５ａ～ｅは個々の語である。図示される実施形態では、各変換３０は、トレーニングシーケンス３２からのトークンのセットを代替トークンと置き換えることを含む。各変換３０は、トークン置換えの目標とされるトレーニングシーケンス３２内の位置の別個のセットを示す、図示されるような別個のバイナリマスクによって定義され得る。図示される例では、Ｔ２マスクをシーケンス３２に適用することは、シーケンス３２の第３および第５のトークンをマスクし、実質的に前記トークンを置換えのためにマークする。

【００２５】

[0035]いくつかの実施形態では、入力修正器４０は、代替トークンのセットを出力して、シーケンス３２内のマスクされたトークンを置き換えるように構成されたトークンジ

ェネレータ４１をさらに備える。図示される例では、トークンジェネレータは、代替トークン３５ｆおよび３５ｇを出力して、それぞれトークン３５ｃおよび３５ｅを置き換える。ジェネレータ４１の単純な実施形態は、基準プールから代替トークンをランダムに引き出すように構成され得る。より高度な実施形態では、トークンジェネレータ４１は辞書／シソーラスを備え、それぞれのマスクされたトークンについて、それぞれのトークンの同意語または反意語を出力するように構成され得る。別の例示的実施形態では、ジェネレータ４１は、マスクされたトークンに従って、さらにはマスクされたトークンのコンテキストに従って代替トークンを決定し得、コンテキストは、マスクされたトークンに先行するシーケンス３２のフラグメント、および／またはマスクされたトークンの後のシーケンス３２のフラグメントからなる。たとえば、図４の例では、トークン３５ｃ（「ｒｅａｄｙ」）のコンテキストは、トークン３５ｂおよび３５ｄ（「ｗｅｒｅ　ｔｏ」）を含み得る。そのような実施形態は、統計言語モデルを利用して、マスクされたトークンのコンテキスト内のそれぞれの代替トークンの発生の確率に従って代替トークンを生成し得る。言い換えれば、トークンジェネレータ４１は、それぞれのトークンシーケンスのコンテキストが与えられると、妥当な代替トークンを生成するように構成され得る。

【００２６】

[0036]妥当な代替トークンを生成するトークンジェネレータ４１の例示的一実施形態は、それぞれの異常検出アプリケーションを表すトークンシーケンスのコーパスに関してトレーニングされたＡＩシステム（たとえば、ディープニューラルネットワークのセット）を備える。ジェネレータ４１のそのようなバージョンは、シーケンス３２内のマスクされたトークンに先行するトークンのサブシーケンスに従って代替トークンを出力し得る。図示される例では、トレーニング済みジェネレータ４１は、トークン３５ａ～ｄのシーケンス（「ｔｈｅｙ　ｗｅｒｅ　ｒｅａｄｙ　ｔｏ」）に続く高い可能性を有するものとして代替トークン３５ｇ（「ｌｅａｖｅ」）を生成することができる。そのようなＡＩベースのトークンジェネレータ４１の一例は、たとえばＪ．Ｄｅｖｌｉｎ他、「ＢＥＲＴ：Ｐｒｅ－ｔｒａｉｎｉｎｇ　ｏｆ　Ｄｅｅｐ　Ｂｉｄｉｒｅｃｔｉｏｎａｌ　Ｔｒａｎｓｆｏｒｍｅｒｓ　ｆｏｒ　Ｌａｎｇｕａｇｅ　Ｕｎｄｅｒｓｔａｎｄｉｎｇ（ＢＥＲＴ：言語理解のためのディープ双方向トランスフォーマの事前トレーニング）」、ａｒＸｉｖ：１８１０．０４８０５に記載されているような、言語のＢｉｄｉｒｅｃｔｉｏｎａｌ　Ｅｎｃｏｄｅｒ　Ｒｅｐｒｅｓｅｎｔａｔｉｏｎ　Ｆｒｏｍ　Ｔｒａｎｓｆｏｒｍｅｒｓ（ＢＥＲＴ）モデルを実装する。いくつかの実施形態では、トレーニング中に、図３～４に示されるようにトークンジェネレータ４１がシーケンスアナライザ４２に接続され、ジェネレータ４１が、シーケンスアナライザ４２の調節可能パラメータとは別個の調節可能な機能パラメータ（たとえば、シナプス重みなど）のセットを有し、トレーニング済みジェネレータ４１の少なくとも１つの調節可能パラメータが、シーケンスアナライザ４２の出力に従って調整されるという意味で、トークンジェネレータ４１は、シーケンスアナライザ４２と共にトレーニングされる。

【００２７】

[0037]入力修正器４０によって実装される別の例示的シーケンス変換３０は、トークン埋込みベクトルの操作を含み得る。そのような実施形態では、修正後シーケンス３４は、トークンのシーケンス自体ではなく、埋込みベクトルの配列を含み得る。入力修正器４０は、当技術分野で一般に埋込み空間と呼ばれる抽象多次元ベクトル空間内のトレーニングシーケンス３２の各トークンの位置を示す座標のセットを決定するように構成されたトークンエンコーダを含み得る。座標のそれぞれのセットは、それぞれのトークンに関連するトークン埋込みベクトルを集合的に定義する。図５は、例示的トークン埋込み空間５０と、トークン３５ａ～ｂをそれぞれ表すトークン埋込みベクトル５５ａ～ｂのセットとを示す。

【００２８】

[0038]例示的埋込み空間は軸のセットによって張られ、各軸は、（たとえば、主成分／特異値分解実施形態において）別個のトークン特徴、またはトークン特徴の一次結合を

表す。コンピューティングイベントのシーケンス中の異常を検出するように構成された実施形態では、トークン特徴は、各イベントの様々なイベント特徴（たとえば、イベントのタイプ、経路標識、ネットワークアドレスなど）を含み得る。好ましい実施形態では、トークンがトレーニングシーケンス内のその位置に従って埋め込まれ、言い換えれば、そのコンテキストに従って埋め込まれる。そのようなケースでは、埋込み空間５０は、抽象コンテキスト空間を含み得、類似のコンテキストで主に生じる２つのトークンが、相対的に近くに共に配置される。とりわけｗｏｒｄ２ｖｅｃ、ＧｌｏＶｅ、およびＢＥＲＴを含む、いくつかのそのような埋込みは当技術分野で周知である。埋込みベクトル表現５５ａ～ｂを生成するために、トークンエンコーダは、トークンシーケンスのコーパスに関してトレーニングされなければならず、トークンシーケンスのコーパスは、トレーニングコーパス１８、すなわちシーケンスアナライザ４２をトレーニングするために使用されるコーパスと一致し得る。トレーニングは、当技術分野で周知の任意の方法に従って、たとえばｂａｇ－ｏｆ－ｗｏｒｄｓおよび／またはｓｋｉｐ－ｇｒａｍアルゴリズムに従って進行し得る。いくつかの実施形態では、トークンエンコーダの調節可能パラメータがシーケンスアナライザ４２の出力に従って調整されるという意味で、トークンエンコーダはアナライザ４２と共にトレーニングされる。

【００２９】

[0039]図６に示されるように、トレーニングシーケンス３２をひずませるためのいくつかのシーケンス変換３０（図３）は、埋込みベクトルに対して直接的に作用し得る。例示的変換T_jは、トレーニングシーケンス３２の選択されたトークンを表す元のトークン埋込みベクトル５５ｃを修正後ベクトル５５ｄに変更し得る。例示的埋込み変換は、軸のうちの１つに沿って、または変換特有の所定の方向に沿って、小さい量εだけベクトルをナッジすることを含む。別の例示的変換は、所定の平面の周りの回転および反射を含み得る。それぞれの変換は、トレーニングシーケンス３２のすべてのトークンに適用され、あるいはたとえば（上記で示したような）バイナリマスクまたは他の選択基準によって識別される、選択されたトークンのみに適用され得る。

【００３０】

[0040]いくつかの実施形態では、シーケンスアナライザ４２は、入力トークンシーケンスを処理して、入力トークンシーケンスに従って決定された変換予測標識３６およびトークン予測標識３８を含む予測標識のセットを生成するように構成される。変換予測標識３６は、入力トークンシーケンスを生成するためにどのシーケンス変換３０が使用された可能性が高いかを示す。例示的実施形態では、変換予測標識３６は、複数の数値スコアＰ（T_1）、Ｐ（T_2）、．．．Ｐ（T_k）を含み、各スコアＰ（T_j）は、それぞれの入力トークンシーケンスを生成するためにそれぞれの変換T_jが適用された可能性を示す。たとえば、標識３６は、入力修正器４０によって実装されるそれぞれの別個のシーケンス変換３０について別個のスコアを含み得る。スコアＰ（T_j）は、所定の境界の間（たとえば、０から１の間）でスケーリングされ得、大きい値は高い可能性を示す。

【００３１】

[0041]次に、トークン予測標識３８のいくつかの実施形態は、入力シーケンスのどのトークンが入力修正器４０によって修正された可能性が高いかを示す。例示的実施形態では、トークン予測標識３８は複数の数値スコアS_1、S_2、．．．を含み、スコアS_nは、入力シーケンスのｎ番目のトークンが入力修正器４０によって変更された可能性を示す。図３に示されるようなトレーニングプロセスでは、各スコアS_nは、修正後トークンシーケンス３４のｎ番目のトークンがそれぞれのトレーニングシーケンス３２のｎ番目のトークンとは異なる可能性を示し得る。

【００３２】

[0042]直感的な視点からは、変換予測標識３６は、トレーニングシーケンス３２をひずませるためにどの変換が使用されたかを推測する試みを表し、トークン予測標識３８は、個々のどのトークンが破壊されたかを推測する試みを表す。標識３６および３８は冗長な情報を伝達するように見えるが（結局のところ、各変換は特定のトークンマスクを有す

る）、標識３６および３８は、シーケンスアナライザ４２の別個のサブシステム（たとえば、ディープニューラルネットワークのニューロンの別個のグループ）によって生成されるという意味で無関係である。さらに、シーケンス変換３０とその関連するトークンマスクとの間の結び付きまたは相関の、シーケンスアナライザ４２内に構築された事前の知識はない。その代わりに、アナライザ４２は、トレーニング中に自動的にそのような相関を学習し得る。いくつかの実施形態は、標識３６と３８の両方を使用すると、シーケンスアナライザ４２のトレーニングをかなり促進し得、たとえば学習を加速し、標識３６および３８のうちのただ１つを使用するのと同様の異常検出性能達成するためにかなり小さいトレーニングコーパスの使用を可能にするという観測に依拠する。

【００３３】

[0043]シーケンスアナライザ３２の例示的アーキテクチャが図７に示されており、層／ニューラルネットワークモジュールのスタックを備え、各層は、前の層／モジュールの出力を受け取り、入力をスタックの次の層に与える。シーケンスアナライザ３２は、トークン表現４８の配列の形で入力を受け取り得、各トークン表現は、入力シーケンスのそれぞれのトークンを特徴付ける数のベクトルを含む。トレーニングシナリオでは、図７の各トークン表現４８は、修正後シーケンス３４（図３）の別個のトークンを表し得る。当技術分野で１ホット符号化と呼ばれるものを用いる例示的一実施形態では、トークン表現４８はＮ×１ベクトルを含み、各行は別個のトークンタイプを表し、Ｎはトークン語彙のサイズを表し、非ゼロ要素は、それぞれのトークンがそれぞれのトークンタイプであることを示す。トークン語彙の例示的サイズＮは、特定の適用分野では数百から数百万の範囲である。別の例では、各トークン表現４８は、当技術分野で周知の埋込みアルゴリズムに従って生成されるトークン埋込みベクトルを含み得る。たとえば、上記で論じられた図５の埋込みベクトル５５ａ～ｂを参照されたい。

【００３４】

[0044]それぞれの連続する層L_iは、それぞれの層に特有のパラメータ（たとえば、活性化、重み、バイアス）のセットに従って、前の層から受け取った入力を変換し、内部ベクトル４９を生成し、そのサイズおよび値の範囲は、アナライザ３２の別個の層／モジュールの間で様々であり得る。たとえば、いくつかの層は、プーリングまたは損失層のケースと同様に、それぞれの入力ベクトルの次元削減を達成する。各層のタイプおよびアーキテクチャは、各実施形態にわたって異なり得る。

【００３５】

[0045]シーケンスアナライザ４２の例示的一アーキテクチャは、畳み込みニューラルネットワーク（ＣＮＮ）層と、その後に続く、整流器（たとえば、ＲｅＬＵまたは他の活性化関数）および／または損失層にさらに結合された密層（すなわち、全結合層）とを含む。代替実施形態は、再帰型ニューラルネットワーク（ＲＮＮ）内に供給されるＣＮＮ層と、その後に続く全結合層およびＲｅＬＵ／損失層とを含み得る。畳込み層は、効率的に、内部ベクトル４９を、当技術分野でフィルタと呼ばれる重みの行列と乗算し、埋込みテンソルを生成し、その結果、それぞれのテンソルの各要素は、選択されたトークンからの寄与を有するが、選択されたトークンに隣接する別のトークンからの寄与も有する。したがって、埋込みテンソルは、個々のトークンよりも粗いグラニュラリティで、入力トークンシーケンスを集合的に表す。フィルタ重みは、トレーニングプロセス中に調節され得る調節可能パラメータである。

【００３６】

[0046]再帰型ニューラルネットワーク（ＲＮＮ）は、人工ニューラルネットワークの特別なクラスを形成し、ネットワークノード間の結び付きが有向グラフを形成する。とりわけ長・短期記憶（ＬＳＴＭ）ネットワークおよびグラフニューラルネットワーク（ＧＮＮ）を含むＲＮＮのいくつかのフレーバが当技術分野で周知である。典型的なＲＮＮは隠れユニット（たとえば、個々のニューロン）のセットを含み、ネットワークのトポロジは、各隠れユニットがそれぞれのトークンm_jを特徴付ける入力（たとえば、埋込みベクトル）だけでなく、隣接する隠れユニットによって与えられる入力も受け取るように具体的

に構成され、隣接する隠れユニットは、入力トークンシーケンス内のトークンm_jに先行するトークンm_{j-1}を特徴付ける入力を受け取る。その結果、各隠れユニットの出力は、それぞれのトークンm_jによって影響を受けるだけでなく、先行するトークンm_{j-1}によっても影響を受ける。言い換えれば、ＲＮＮ層は、前のトークンのコンテキスト内の各トークンについての情報を処理し得る。双方向ＲＮＮアーキテクチャは、入力トークンシーケンスの前のトークンと後続のトークンの両方のコンテキスト内の各トークンについての情報を処理し得る。

【００３７】

[0047]シーケンスアナライザ４２のさらに別の例示的実施形態は、トランスフォーマニューラルネットワーク層のスタックを備え得る。トランスフォーマアーキテクチャは、とりわけ、たとえばＡ．Ｖａｓｗａｎｉ等「Ａｔｔｅｎｔｉｏｎ ｉｓ ａｌｌ ｙｏｕ ｎｅｅｄ（あなたが必要なすべてのものは注意である）」、ａｒＸｉｖ：１７０６．０３７６２に記載されている。各入力トークンシーケンスについて、トランスフォーマ層は、コンテキスト化されたトークン埋込みベクトルのシーケンスを生成し得、各トークン埋込みベクトルh_jは、入力シーケンスの複数の（たとえば、すべての）トークンm_jからの情報を符号化する。トランスフォーマ層の出力は、当技術分野で予測ヘッドと呼ばれ、図７のブロックＨ１およびＨ２として示される複数の別個の分類器モジュール（たとえば、密層）内に供給され得る。次いで、ヘッドＨ１およびＨ２は、それぞれ変換予測標識３６およびトークン予測標識３８を出力し得る。

【００３８】

[0048]図８は、本発明のいくつかの実施形態による、異常検出器２０をトレーニングするためにＡＩトレーニングアプライアンス１４（図１）によって実施されるステップの例示的シーケンスを示す。コーパス１８からトレーニングシーケンス３２を選択することに応答して、ステップ２０４～２０６のシーケンスは、シーケンス変換３０の利用可能なセットから変換を（たとえば、ランダムに）選択し、それぞれの変換をトレーニングシーケンス３２に適用し、したがって修正後トークンシーケンス３４を生成し得る。

【００３９】

[0049]修正後シーケンス３４はシーケンスアナライザ４２に供給され、シーケンスアナライザ４２は、修正後シーケンス３４を処理して、予測標識３６および／または３８を生成する。次いで、ステップ２０８は、トレーニングシーケンス３２および／または予測標識３６および／または３８に従ってユーティリティ関数を決定し得る。機械学習の技術分野で損失とも呼ばれる例示的ユーティリティ関数は、以下のように表現され得る。

【００４０】

$$\Sigma_{\chi} U(\theta, \chi) \quad [1]$$

上式でχはトレーニングシーケンスを表し、θは調節可能パラメータのセットを表す。トレーニングは、ユーティリティＵを最小にする方向にパラメータθを調節することを含み得る。

【００４１】

[0050]単純なユーティリティ関数は、所望の出力からのアナライザ４２の出力の逸脱を定量化し得る。たとえば、例示的ユーティリティ関数は、ステップ２０６でどの変換が適用されたか、および／またはステップ２０６で元のトレーニングシーケンス３２のどのトークンが破壊されたかをアナライザ４２が正しく推測したかどうかを示し得、誤った推測に対してアナライザ４２にペナルティを課し得る。

【００４２】

[0051]いくつかの実施形態は、標識３６および３８に従って決定されたユーティリティを組み合わせると、トレーニングを促進し、かつ／またはより高性能のシーケンスアナライザ４２が得られ得るという観測に依拠する。好ましい実施形態は、シーケンスレベル成分（シーケンス変換の選択を示す）をトークンレベル成分（それぞれの個々のトークンが破壊されたか否かを示す）と組み合わせるアグリゲートユーティリティ関数を使用し得る。

【００４３】
$U = \lambda_1 U_S + \lambda_2 U_T$ ［２］
上式でU_SおよびU_Tは、それぞれシーケンスレベル成分およびトークンレベル成分を表し、λ_1およびλ_2は、各ユーティリティ関数の相対的寄与を変更することを可能にする重みである。最尤トレーニング方法を実装するいくつかの実施形態では、
【００４４】
【数１】

$$U_S = \mathbb{E}_{k,x}[-\log P(T_k|\tilde{x}, \theta_A)], \qquad [3]$$

【００４５】
上式で
【００４６】
【数２】

$$\mathbb{E}$$

【００４７】
は期待値を表し、
【００４８】
【数３】

$$P(T_k|\tilde{x}, \theta_A)$$

【００４９】
は、ひずんだシーケンス
【００５０】
【数４】

$$\tilde{x}$$

【００５１】
がシーケンス変換T_Kの適用によって生成された確率を表し（たとえば、図３の変換予測標識３６を参照）、θ_Aは一般に、シーケンスアナライザ４２の調節可能パラメータを表す。一方、
【００５２】
【数５】

$$U_T = \mathbb{E}_{k,x}[\sum_i -\log S_i\ (T_k|\tilde{x},\ \theta_A)], \quad [4]$$

【００５３】
上式で
【００５４】
【数６】

$$S_i(T_k|\tilde{x},\ \theta_A)$$

【００５５】
は、トレーニングシーケンスＸのトークンｉがシーケンス変換T_Kの適用によって影響を受けた確率を表す（たとえば、図３の変換予測標識３８を参照）。
[0052]いくつかの実施形態では、入力修正器４０の構成要素（トークン埋込みベクトルを生成するように構成されたトークンジェネレータ４１および／またはトークンエンコーダなど）が、シーケンスアナライザ４２と共にトレーニングされる。そのような実施形態は、前述のU_SおよびU_Tに加えて、ジェネレータユーティリティ関数を使用し得る。
【００５６】
$U = \lambda_1 U_S + \lambda_2 U_T + \lambda_3 U_G$ ［５］
上式で、λ_3は、ジェネレータユーティリティ関数U_Gのグローバルユーティリティへの寄与を調節するために使用される別の重みを表し、
【００５７】
【数７】

$$U_G = \mathbb{E}_{k,x}[\sum_i -\log P_G\ (t_i|\tilde{x}, \theta_G)], \quad [6]$$

【００５８】
上式でθ_Gは一般に、トークンジェネレータ４１の調節可能パラメータを表し、
【００５９】
【数８】

$$P_G\ (t_i|\tilde{x}, \theta_G)$$

【００６０】
は、トークンt_iが修正後シーケンス３４中に現れる確率、または言い換えれば、トークンt_iが修正後シーケンス３４のコンテキスト内で妥当である確率を表す。
[0053]次いで別のステップ２１０は、決定されたユーティリティ関数に従って、パラメータθ_Aおよび／またはθ_Gのセットを調節し得る。そのような調節は、勾配降下によ

る逆伝播手順、または選ばれたユーティリティ関数を最小にすることを目標とする任意の他の最尤探索を実装し得る。トレーニングは、終了条件が満たされるまで、たとえば所定のエポック数について、所定の数のトレーニングシーケンスが解析されるまで、所定のレベルの異常検出性能が実証されるまでなどの間、続行し得る（ステップ２１２）。トレーニングの成功に応答して、シーケンスアナライザ４２の調節可能パラメータ（たとえば、シナプス重みなど）の最適な値が、検出器パラメータ値２４（図２）の形でエクスポートされ、クライアントシステム１０ａ～ｃに送られる。

【００６１】

[0054]図９は、本発明のいくつかの実施形態による、異常を検出するためにクライアントシステム１０ａ～ｃおよび／またはユーティリティサーバ１２によって実施されるステップの例示的シーケンスを示す。ステップ２２２～２２４のシーケンスは、ＡＩトレーニングアプライアンス１４（図２）から検出器パラメータ値２４を受け取り、それぞれの値で検出器２０のローカルインスタンスをインスタンス化することによって、動作のために検出器２０を準備する。次いで、各ターゲット・トークン・シーケンス２２について、ステップ２２８は、シーケンスアナライザ４２を実行して、それぞれのターゲット・トークン・シーケンス２２についてのトークン予測標識３８を決定し得る。

【００６２】

[0055]別のステップ２３０は、判断モジュール４４を適用して異常標識２６を生成し得る。いくつかの実施形態では、判断モジュール４４は、たとえばターゲット・トークン・シーケンス２２全体にわたって取られた個々のトークン予測スコアS_iの平均として、トークン予測標識３８に従って異常標識２６を決定するように構成される。それぞれの個々のスコアS_iは、図３に関連して上記で説明されたように、シーケンス２２のそれぞれのトークンがシーケンス変換３０の適用によって破壊されたかどうかの可能性を定量化し得る。いくつかの実施形態は、以下に従ってシーケンス特有の異常スコアＡを決定し得る。

【００６３】

【数９】

$$A = \frac{1}{L_S}\sum_i S_i \,, \qquad [7]$$

【００６４】

[0056]上式で、L_Sは、ターゲットシーケンス２２の長さ（トークンのカウント）を表す。大きいS_iが、それぞれのトークンが破壊された可能性が高いことを示す一実施形態では、Ａの大きい値は、ターゲット・トークン・シーケンス２２が異常である可能性が高いことを示し得る。逆に、大きいS_iが、それぞれのトークンが破壊されていない可能性が高いことを示すとき、大きいＡの値は、ターゲットシーケンス２２が異常ではないことを示し得る。判断モジュール４４のいくつかの実施形態は、異常スコアＡの計算された値を所定のしきい値と比較し、比較の結果に従ってターゲット・トークン・シーケンス２２が異常であるか否かを判定する。

【００６５】

[0057]図１０は、本明細書で説明される方法のうちのいくつかを実行するようにプログラムされたコンピューティングアプライアンス７０の例示的ハードウェア構成を示す。コンピューティングアプライアンス７０は、図１のクライアントシステム１０ａ～ｃ、ユーティリティサーバ１２、およびＡＩトレーニングアプライアンス１４のいずれかを表し得る。図示されるコンピューティングアプライアンスはパーソナルコンピュータであり、

サーバ、携帯電話、タブレットコンピュータ、ウェアラブルなどの他のデバイスは、わずかに異なる構成を有し得る。プロセッサ７２は、信号および／またはデータのセットと共に計算および／または論理演算を実行するように構成された物理デバイス（たとえばマイクロプロセッサ、半導体基板上に形成されたマルチコア集積回路）を備え得る。そのような信号またはデータは、プロセッサ命令、たとえば機械コードの形で符号化され、プロセッサ７２に送達され得る。

【００６６】

[0058]プロセッサ７２は一般に命令セットアーキテクチャ（ＩＳＡ）によって特徴付けられ、ＩＳＡは、とりわけ、プロセッサ命令のそれぞれのセット（たとえばｘ８６ファミリとＡＲＭ（登録商標）ファミリ）、およびレジスタのサイズ（たとえば、３２ビットプロセッサと６４ビットプロセッサ）を指定する。プロセッサ７２のアーキテクチャは、所期の主な用途に従って様々であり得る。中央演算処理装置（ＣＰＵ）は汎用プロセッサであり、グラフィックス処理装置（ＧＰＵ）は、イメージ／ビデオ処理およびいくつかの形態の並列コンピューティングのために最適化される。プロセッサ７２は、Ｇｏｏｇｌｅ（登録商標），Ｉｎｃ．によるテンソル処理装置（ＴＰＵ）、様々な製造業者によるニューラル処理装置（ＮＰＵ）などの特定用途向け集積回路（ＡＳＩＣ）をさらに含み得る。ＴＰＵおよびＮＰＵは、本明細書で説明されるような機械学習アプリケーションに特に適していることがある。

【００６７】

[0059]メモリユニット７４は、動作を実施する過程でプロセッサ７２によってアクセスされ、または生成されるデータ／信号／命令符号化を記憶する揮発性コンピュータ可読媒体（たとえば、ダイナミックランダムアクセスアクセスメモリ－ＤＲＡＭ）を備え得る。入力デバイス７６は、とりわけ、ユーザがデータおよび／または命令をアプライアンス７０内に導入することを可能にするそれぞれのハードウェアインターフェースおよび／またはアダプタを含む、コンピュータキーボード、マウス、およびマイクロフォンを含み得る。出力デバイス７８は、とりわけモニタやスピーカなどのディスプレイデバイス、ならびにそれぞれのコンピューティングアプライアンスがユーザにデータを通信することを可能にする、グラフィックカードなどのハードウェアインターフェース／アダプタを含み得る。いくつかの実施形態では、入力および出力デバイス７６～７８は、共通のハードウェア（たとえば、タッチスクリーン）を共有する。記憶デバイス８２は、ソフトウェア命令および／またはデータの不揮発性記憶、読取り、および書込みを可能にするコンピュータ可読媒体を含む。例示的記憶デバイスは、磁気ディスク、光ディスク、およびフラッシュメモリデバイス、ならびにＣＤおよび／またはＤＶＤディスクおよびドライブなどの取外し可能媒体を含む。ネットワークアダプタ８４は、コンピューティングアプライアンス７０が電子通信ネットワーク（たとえば、図１のネットワーク１５）および／または他のデバイス／コンピュータシステムに接続することを可能にする。

【００６８】

[0060]コントローラハブ８０は一般に、複数のシステム、周辺機器、および／またはチップセットバス、ならびに／あるいはプロセッサ７２とアプライアンス７０のハードウェア構成要素の残りの部分との間の通信を可能にするすべての他の回路を表す。たとえば、コントローラハブ８０は、メモリコントローラ、入力／出力（Ｉ／Ｏ）コントローラ、および割込みコントローラを備え得る。ハードウェア製造業者に応じて、いくつかのそのようなコントローラは単一の集積回路内に組み込まれ得、かつ／またはプロセッサ７２と一体化され得る。別の例では、コントローラハブ８０は、プロセッサ７２をメモリ７４に接続するノースブリッジ、ならびに／あるいはプロセッサ７２をデバイス７６、７８、８２、および８４に接続するサウスブリッジを備え得る。

【００６９】

[0061]前述の例示的システムおよび方法は、様々な適用分野で異常の効率的な自動検出を可能にする。いくつかの実施形態では、トレーニングコーパスから引き出されるトークンシーケンスが、トレーニングを施すシーケンスアナライザに供給される前に、複数の

所定のシーケンス変換のうちの少なくとも１つに従ってひずめられる。次いでシーケンスアナライザは、それぞれの入力トークンシーケンスを生成するためにどの変換が使用されたかを正しく推測するようにトレーニングされる。

【００７０】

[0062]異常検出器をトレーニングするためのいくつかの従来の手順は、トークンのうちのいくつかをランダムに置き換えることによってトレーニング・トークン・シーケンスを破壊し、その後で、どのトークンが置き換えられたかを推測するように検出器をトレーニングする。しかしながら、そのようなトレーニング方法は、計算資源の点で比較的コストがかかり、トレーニングコーパスおよび／またはアプリケーションのいくつかの選択肢については不安定であり得る。この従来の手法とは対照的に、本発明のいくつかの実施形態は、変換の事前定義されたセットを使用して入力トークンシーケンスをひずませ、トークンレベル成分（それぞれの個々のトークンが破壊されたか否かを示す）をシーケンスレベル成分（入力シーケンス全体をひずませる方式を示す）と組み合わせるアグリゲートユーティリティ関数に従ってトレーニングする。より従来型のトークンレベルユーティリティに加えてシーケンスレベルユーティリティを使用することは、反直感的に見えることがある。いくつかの実施形態では、各シーケンス変換が特定のトークンマスクを有し、したがってどの変換が適用されたかを推測することが、実質的にはどのトークンが破壊されたかも推測することになり得るからである。しかしながら、いくつかの実施形態は、シーケンスレベルの学習タスク（適用された変換を推測すること）をセットアップするのと同時に、トークンレベルのタスク（特定のトークンが破壊されたかどうかを推測すること）をセットアップすることは、シーケンスアナライザが存在しないマスクパターンを予測するのを阻止することによって、正しい学習を強化し得る。いくつかの実施形態では、トークンレベルの予測およびシーケンスレベルの予測が、ディープニューラルネットワークの別個の予測ヘッドによって生成される。したがって、異常検出器内に構築される変換とトークンマスクとの間の相関の事前の知識はない。その代わりに、検出器は、トレーニング中にそのような潜在的相関を学習し、そのことは、よりロバストなモデルに至り得る。

【００７１】

[0063]コンピュータ実験は、トークンレベルのタスクとシーケンスレベルのタスクを組み合わせると、検出器の性能を改善することによって学習を促進することを示している。逆に、本明細書で説明されるようなトレーニング方法を使用することにより、より小さいトレーニングコーパスおよび／またはより少ないネットワークパラメータを使用して、同一のレベルの異常検出性能が達成され得る。これは、トレーニングコーパス１８が比較的小さいサイズを有する状況（たとえば、トレーニングコーパスがソーシャルメディアポストからなるとき）での作成者属性などの異常検出タスクにとって特に有利であり得る。本明細書で説明されるように検出器をトレーニングすることは、直感的にはトレーニングコーパスのサイズを人工的に増加させることに対応する。同一のトレーニング・トークン・シーケンス３２が、別個のシーケンス変換３０の適用に応答して複数の別個の修正後シーケンス３４を引き起こし得るからである（図３参照）。

【００７２】

[0064]いくつかの実施形態は、第２のＡＩシステムを利用して、トレーニングシーケンスの妥当なひずみを生成する。たとえば、それぞれのトークンシーケンスの残りの部分のコンテキストが与えられると、ＢＥＲＴ言語モデルを実装するトークンジェネレータが使用され、選択されたトークンが妥当な置換で置き換えられ得る。いくつかの実施形態は、既にトレーニングされたジェネレータが、異常と見なされるにはある意味で「妥当過ぎる」修正後トレーニングシーケンスを生成することによって学習を妨げ得るという観測に依拠して、事前トレーニング済みの高性能バージョンのトークンジェネレータを使用する代わりに、トークンジェネレータを異常検出器と共に明示的にトレーニングする。共にトレーニングすることは、シーケンスアナライザが修正を検出することにより熟練するにつれて、トークンジェネレータが妥当な修正後トレーニングシーケンスを生成することにますます熟練することを保証し得る。さらに、異常検出器をトレーニングするために使用さ

れるコーパスとは別個のコーパスに関してトークンジェネレータを事前トレーニングすることは、異常値情報をもたらし、異常検出器が異常値情報をそのように認識することを防止し得る。

【００７３】

[0065]本発明のいくつかの実施形態に従ってトレーニングされた異常検出器は、とりわけ、以下を含む様々なシナリオで使用され得る。

自動テキスト分類

[0066]例示的自然言語処理（ＮＬＰ）アプリケーションでは、異常検出器２０は、特定のカテゴリ（たとえば、ビジネスニュース）に属するテキストからなるコーパスに関してトレーニングされ、次いでターゲット・テキスト・フラグメントがそれぞれのカテゴリに属するか否かを判定するために使用され得る。そのような実施形態では、高い異常スコアは、それぞれのテキストがそれぞれのカテゴリに属さないことを示し得る。

【００７４】

[0067]コンピュータ実験では、本明細書で説明される異常検出器が、ニュース記事の標準基準コーパス（２０Ｎｅｗｓｇｒｏｕｐｓ）のサブセットに関してトレーニングされ、サブセットは、選択されたカテゴリ（コンピューティング、リクリエーション、科学、雑報、政治、または宗教）の記事からなる。実験は、４つの積重ねトランスフォーマ層を含み、２つの予測ヘッドが上端にあるシーケンスアナライザを使用した。各トランスフォーマ層は、４つの自己注意ヘッド、サイズ２５６の隠れ層、ならびにサイズ１０２４および２５６のフィードフォワード層を含んだ。各予測ヘッドは、非線形性によって分離される２つの線形層を有し、分類層で終了した。トレーニング・トークン・シーケンスの最大サイズは１２８であった。シーケンス変換３０は、ランダムトークンジェネレータを使用して、別個のマスクパターンに従ってトークンを置き換えることからなるものであった。様々なカウントおよびカバレッジのマスクパターンが、入力トレーニングシーケンスの２５％から５０％の間をカバーする、５から１００個の間の別個のマスクパターンと共に試行された。

【００７５】

[0068]次いで、トレーニングされた検出器は、集合からランダムに選択された記事が、検出器がトレーニングされたカテゴリに属するかどうかを識別するように求められた。本発明のいくつかの実施形態に従ってトレーニングされた異常検出器は、それぞれのタスクで現況技術の従来の異常検出器より一貫して大幅に優れており、典型的なａｒｅａ ｕｎｄｅｒ ｔｈｅ ｒｅｃｅｉｖｅｒ ｏｐｅｒａｔｉｎｇ ｃｕｒｖｅ（ＡＵＲＯＣ）値は、約７０％（科学カテゴリに関してトレーニングされたとき）から９２％超（コンピューティングニュースに関してトレーニングされたとき）までの範囲である。一般には、別個の変換数を増加させると、トレーニング済みの異常検出器の性能を改善し、トークン埋込み内の表現性を促進することを実験は明らかにした。２５％から５０％の割合の破壊されたトークンを有する変換が最良の結果を生み出すように見えた。

自動作成者属性

[0069]異常検出器２０のいくつかの実施形態は、選択された作成者によって書かれたテキスト（たとえば、手紙、記事、ブログポスト、ｅメール、ソーシャルメディアポスト）のコーパスに関してトレーニングされ、次いで、ターゲット・テキスト・フラグメントがそれぞれの人によって作成されたかどうかを判定するために使用され得る。例示的アプリケーションは、匿名の手紙の作成者を判定し、様々な文書の真正性、および文学作品の死後の属性を検証することを含む。いくつかの実施形態はまた、犯罪学の適用分野をも有し得る。たとえば、Ｄａｒｋ Ｗｅｂリソースの作成者またはユーザを識別する、たとえば盗品、クレジットカードデータ、小児ポルノ、銃、麻薬などの取引などの犯罪活動に関与するユーザについての集合地点として役目を果たすフォーラム上でポストされたメッセージの作成者を識別する際に法執行に関心が持たれることがある。「Ｄａｒｋ Ｗｅｂ」という用語は、本明細書では、検索エンジンによって索引付けされず、かつ／またはプライベートピアツーピアネットワークもしくはＴｏｒなどの匿名化ソフトウェアを介しての

みアクセス可能であるコンテンツを表すために使用される。
【００７６】
[0070]次いで、容疑者のセットによって作成され、公に利用可能であるオンラインコンテンツ（たとえば、人気のあるソーシャルメディアサイトおよび／またはユーザフォーラム上でそれぞれの容疑者によってポストされたコメント）のコーパスに関してトレーニングされた異常検出器の例示的実施形態が、Ｄａｒｋ　Ｗｅｂから取り入れられたターゲット・テキスト・フラグメントを解析するために使用され得る。ターゲットテキストが異常ではないことを示す異常スコアが、ターゲットテキストの作成者が、検出器がトレーニングされたテキストのコーパスの作成者のうちの１人と一致することを示し得る。
ソーシャルメディア監視
[0071]異常検出器２０の一実施形態が、ソーシャルメディアアカウントの選択されたセットに関連するウェブコンテンツ、たとえばＴｗｉｔｔｅｒ（登録商標）フィードの特定の集合に関してトレーニングされ得る。トレーニングコーパスは、特定の時間ウィンドウ（たとえば、１日、１週など）以内に発行されたコンテンツにさらに限定され得る。次いで、検出器が使用され、新しくポストされたコンテンツが解析され得る。異常は、トピックの変化および／または進行中の交換のトーンの変化を示し得、したがって新しいトピックおよびトレンドの適時の自動検出を可能にする。
フェイクおよび自動生成されたコンテンツの検出
[0072]異常検出器２０の一実施形態は、選択された人間の作成者（たとえば、実際のニュース記事、実際のユーザによってポストされたソーシャルメディア）によって書かれたテキストのコーパスに関してトレーニングされ得る。コーパスは、選択された定期刊行物、新聞、またはニュースウェブサイトについて書かれた記事、または選択されたジャーナリストによって書かれた記事にさらに狭められ得る。次いで、トレーニングされた異常検出器は、ターゲット・テキスト・フラグメントを解析するために使用され得る。ターゲットテキストが異常であることを示す異常スコアは、それぞれのテキストがフェイクニュースを含み得ること、および／またはマシンで生成され得ることを示し得る。
データ保護およびプライバシー
[0073]いくつかのクラウドコンピューティングサービスは、ユーザが他のユーザと共有するために、または様々な操作（たとえば、マルウェア走査）のためにリモートサーバにファイルをアップロードすることを可能にする。一例として、ユーザのコンピュータ上で実行中のソフトウェアエージェントは、それぞれのユーザによって示され得る、選択されたフォルダのコンテンツを自動的にアップロードし得る。クラウドにデータをアップロードすることは、ユーザがアップロード用のコンテンツを明示的に選ばないときは特に、プライバシーリスクを含み得る。たとえば、ユーザが何らかの機密データ（たとえば、個人ファイルまたは写真、医療記録など）をアップロードフォルダに誤ってドロップした場合、それぞれのデータがユーザの希望に反して自動的にアップロードされる。
【００７７】
[0074]異常検出器２０の一実施形態が、ユーザのコンピュータ上にインストールされ、通常はそれぞれのユーザによってアップロードされるファイル、たとえばリモート走査のために最近アップロードされた１００個のファイルに関してトレーニングされ得る。追加のフィルタが、Ｐｏｒｔａｂｌｅ　Ｄｏｃｕｍｅｎｔ　Ｆｏｒｍａｔ（ＰＤＦ）文書やＭｉｃｒｏｓｏｆｔ（登録商標）Ｏｆｆｉｃｅ（登録商標）ファイルなどの特定の種類のファイルのみを選択し得る。そのような実施形態は、前述のような自然言語処理技法を使用し得、トークンは個々の語などを含む。次いで、トレーニング済みの異常検出器が、アップロードに備えてリモート走査のために現在目印が付けられている各ファイルを解析するために使用され得る。それぞれのファイルについて決定された異常スコアが、潜在的異常を示すとき、いくつかの実施形態は、それぞれのファイルがアップロードされるのを防止し得、ユーザに通知し得る。
コンピュータセキュリティ
[0075]異常検出器２０のいくつかの実施形態は、活動の正常パターンを表すと見なさ

れる基準時間間隔中に生じるコンピューティングイベントのシーケンスに関してトレーニングされ、次いでセットクライアントコンピュータシステムの挙動を監視するために使用され得る。クライアント上で検出された異常な挙動は、コンピュータセキュリティ脅威、たとえばそれぞれのクライアントが悪意のあるソフトウェアを実行していること、または侵入者／ハッカーがそれぞれのクライアントへのアクセス権を得たことを示し得る。

【００７８】

[0076]いくつかの実施形態では、異常検出は、監視されるソフトウェアエンティティ（たとえば、プロセス、仮想マシンなど）の実行中に生じるイベントのシーケンスを解析することを含む。そのようなイベントの例は、とりわけ、プロセス／スレッドの起動（たとえば、ユーザがアプリケーションを起動する、親プロセスが子プロセスを生み出すなど）、それぞれのクライアントシステムの入力デバイス（たとえば、カメラ、マイクロフォン）にアクセスする試み、ローカルまたはリモートネットワークリソースにアクセスする試み（たとえば、特定のＵＲＬのアクセスするためのハイパーテキスト転送プロトコル－ＨＴＴＰ要求、ローカルネットワークを介して文書リポジトリにアクセスする試み）、特定のユニフォームリソース識別子方式で構築された要求（たとえば、ｍａｉｌｔｏ：またはＦＴＰ：要求）、特定のプロセッサ命令（たとえば、システムコール）の実行、ライブラリ（たとえば、ダイナミックリンクライブラリ－ＤＬＬ）をロードする試み、新しいディスクファイルを作成する試み、ディスク上の特定の場所から読み取り、または特定の場所に書き込む試み（たとえば、既存のファイルを上書きする試み、特定のフォルダまたは文書を開く試み）、および電子メッセージ（たとえば、ｅメール、ショートメッセージサービス－ＳＭＳなど）を送る試みを含む。いくつかの実施形態では、不活動の期間、イベント間の時間ギャップおよび／またはそれぞれのクライアントシステムが遊休状態であり、ユーザ活動を登録せず、または内部システムタスクのみを実施するときの時間間隔も、イベントとしての資格が与えられ得る。本明細書で説明されるシステムおよび方法が、とりわけソーシャルメディア上のユーザの活動に関するイベント、ユーザのブラウジング履歴、ユーザのゲーミング活動などの他の種類のイベントを解析するように適合され得ることを当業者は理解されよう。

【００７９】

[0077]イベント検出は、当技術分野で周知の任意の方法を含み得る。一例として、保護されたクライアント上で実行中のセキュリティエージェントは、監視されるソフトウェアエンティティのセットを、ｅｖｅｎｔ　ｔｒａｃｋｉｎｇ　ｆｏｒ　Ｗｉｎｄｏｗｓ（登録商標）などのＯＳ４０のイベントロギングサービスに登録し得る。それに応答して、エージェントは、それぞれのプロセスの実行中に生じる様々なイベントの通知を、リアルタイムで、またはログ形式で受信し得る。イベントロギングツールは通常、各イベントについてのタイムスタンプ、イベントタイプを識別する数値コード、それぞれのイベントを生成したプロセスまたはアプリケーションのタイプの標識、および他のイベントパラメータを含むイベント記述子のリストを生成する。イベントシーケンスは、ログを構文解析することによって組み立てられ得る。

【００８０】

[0078]いくつかの実施形態は、各イベントを別々のトークンとして扱い得る。トークンはイベント語彙に従って符号化され得、イベント語彙は、数千から数百万個の別個のイベントタイプを含み得る。次いで、異常検出器をトレーニングすることは、前述のようにトレーニングイベントシーケンスに様々な変換を適用することを含み得る。例示的シーケンス変換は、トレーニングシーケンスの選択されたイベントを削除、挿入、および並べ替えること、ならびに選択されたイベントを異なる種類の代替イベントと置き換えることを含み得る。

【００８１】

[0079]代替実施形態は、イベントログエントリをテキストトークンのシーケンスと見なし得る。たとえば、ログエントリ：

２０：１０｜ＩＮＦＯ｜ｍａｎａｇｅｒ．ｓｔｒａｇｅ｜Ｆｏｕｎｄ　ｂｌｏｃｋ　ｒ

ｄｄ＿２＿３　ｌｏｃａｌｌｙ
が以下のトークンシーケンスに構文解析され得る。
【００８２】
　２０：１０；ＩＮＦＯ；ｍａｎａｇｅｒ；ｓｔｏｒａｇｅ；Ｆｏｕｎｄ；ｂｌｏｃｋ；ｒｄｄ＿２＿３；ｌｏｃａｌｌｙ
　ただし個々のトークンがセミコロンで分離される。次に、入力修正器４０が、選択されたトークンを代替と置き換えることによって、それぞれのトークンシーケンスをひずませ得る。置換のために選択されたトークンの位置が、前述のようにマスクによって示され得る。そのような一例として、上記について決定された修正後トークンシーケンスは以下のように理解され得る。
【００８３】
　２０：１０；ＤＥＢＵＧ；ｍａｎａｇｅｒ；ｔｈｒｅａｄ；Ｆｏｕｎｄ；ｂｌｏｃｋ；ｒｄｄ＿２＿３；ｇｌｏｂａｌｌｙ
　ただし、代替トークンが太字で示されている。いくつかの実施形態では、ブログエントリの選択されたフィールドが修正されないようにマスクが選ばれる。トークンジェネレータ４１は、候補のフィールド特有のプールまたは位置特有のプールから代替トークンを選択するように構成され得る。上記の例では、第２のトークンについての候補代替のプールが｛ＷＡＲＮＩＮＧ，ＤＥＢＵＧ，ＩＮＦＯ，ＣＲＩＴＩＣＡＬ｝からなることがある。
【００８４】
　[0080]ログに対する異常検出の例示的適用は、ハニーポットシステム上に記録されたアクセスおよび／またはイベントログを解析することによってゼロデイ攻撃を検出することを含む。本明細書で説明されるような異常検出器は、ログの第１の部分に関してトレーニングされ得、したがってログの第１の部分に対応する時間枠の間のそれぞれのハニーポットの「正常な」挙動を学習する。次いで、異常検出器は、ログの第２の部分を解析するために使用され得る。異常がログの第１の部分と第２の部分との間のハニーポットの挙動の変化を示し得、新しいマルウェアの可能な出現、ボットネットの活動化などを示唆する。いくつかの実施形態は、異常検出器を周期的に（たとえば、前の時間からのログデータに関して毎時間ごとに）再トレーニングし、それを使用して新しい脅威をリアルタイムに監視し得る。
【００８５】
　[0081]本発明の範囲から逸脱することなく、上記の実施形態が多くの方式で変更され得ることは、当業者にとっては明らかであろう。したがって、本発明の範囲は、以下の特許請求の範囲およびその法的均等物によって決定されるべきである。
【図面】
【図１】

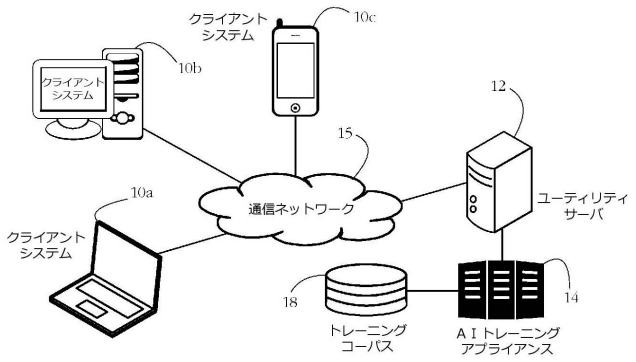

FIG. 1

【図２】

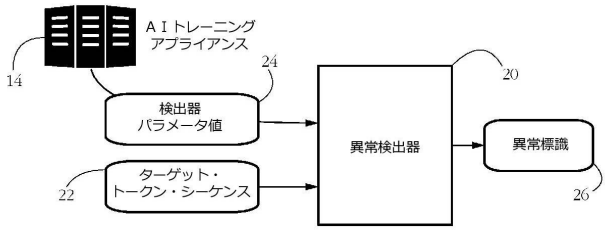

FIG. 2

【図３】

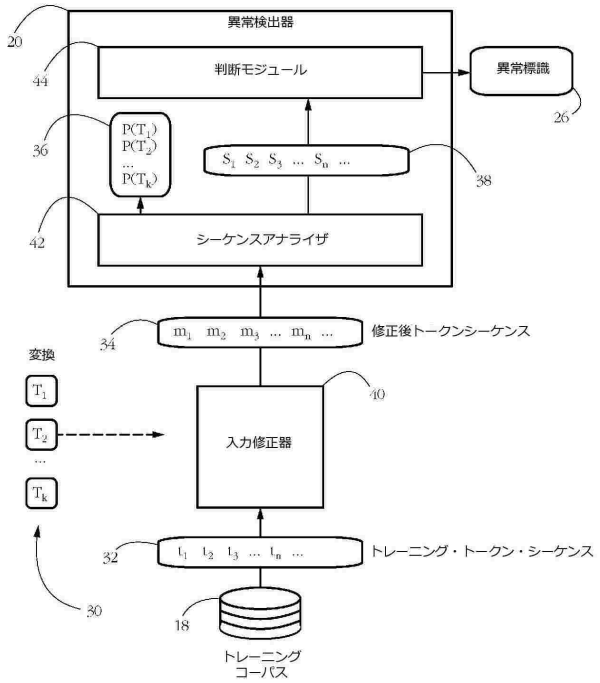

FIG. 3

【図４】

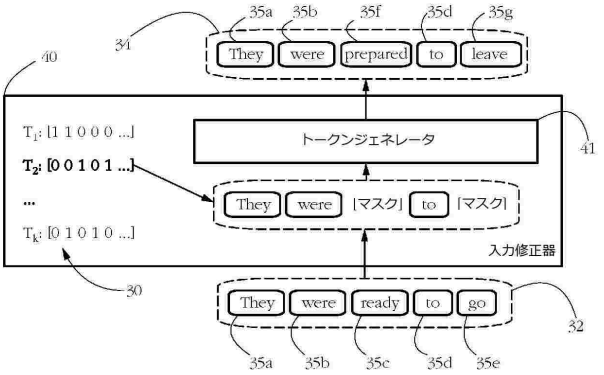

FIG. 4

【図５】

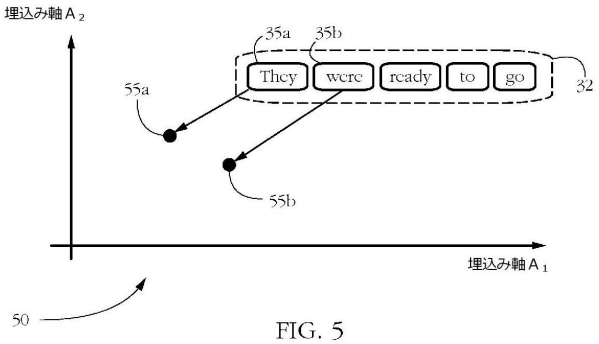

FIG. 5

【図６】

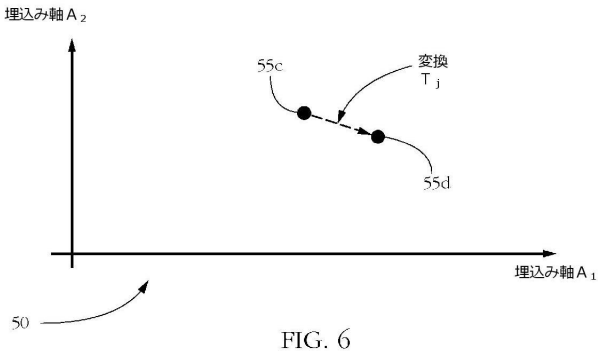

FIG. 6

【図７】

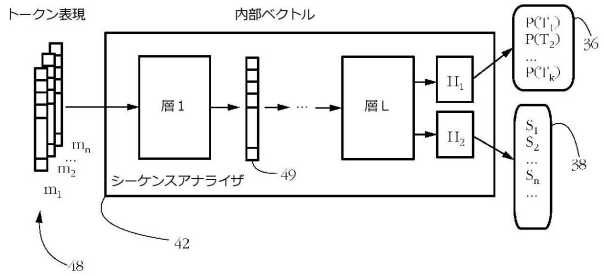

FIG. 7

【図８】

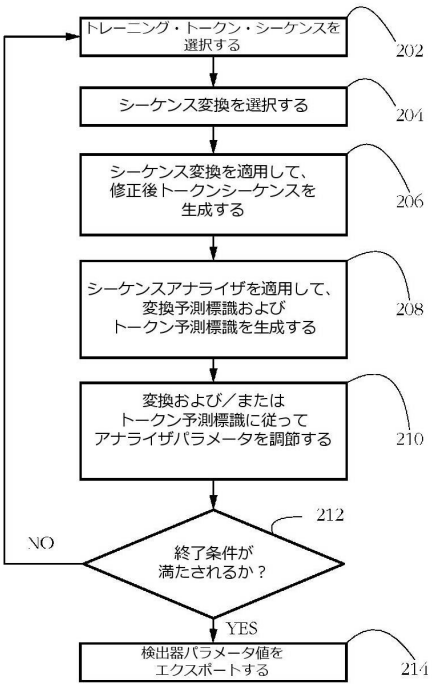

FIG. 8

【図９】

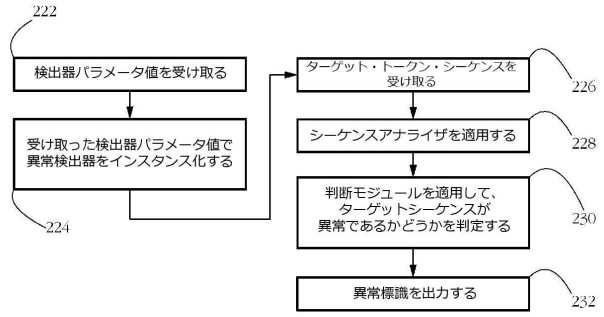

FIG. 9

【図１０】

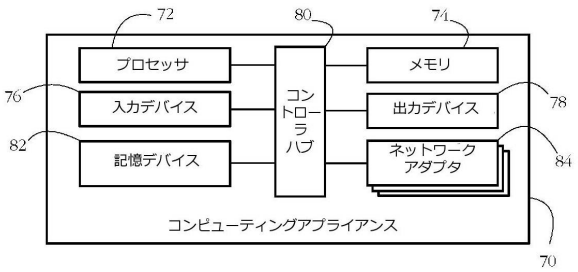

FIG. 10

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2022/058130

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06N3/04 G06N3/08
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2020/255137 A1 (YISSUM RES DEV CO OF HEBREW UNIV JERUSALEM LTD [IL]) 24 December 2020 (2020-12-24) paragraph [0026] - paragraph [0038] paragraph [0057] - paragraph [0073] figure 1 -----	1-21
A	EP 3 726 409 A2 (CROWDSTRIKE INC [US]) 21 October 2020 (2020-10-21) paragraph [0029] - paragraph [0046] paragraph [0166] - paragraph [0176] -----	1-21

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
23 August 2022	08/09/2022

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Baldan, Marco

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2022/058130

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2020255137 A1	24-12-2020	EP 3987455 A1	27-04-2022
		US 2022253699 A1	11-08-2022
		WO 2020255137 A1	24-12-2020
EP 3726409 A2	21-10-2020	EP 3726409 A2	21-10-2020
		US 2020327225 A1	15-10-2020

Form PCT/ISA/210 (patent family annex) (April 2005)

フロントページの続き

MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,CZ,DE,DJ,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IR,IS,IT,JM,JO,JP,KE,KG,KH,KN,KP,KR,KW,KZ,LA,LC,LK,LR,LS,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,SD,SE,SG,SK,SL,ST,SV,SY,TH,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC,VN,WS,ZA,ZM,ZW

トラダ・ミトロポリト・ベニアミン・コスタチェ・ヌマルル １，

(72)発明者 フローリン・エム，ブラッド

ルーマニア国 １０５６００ クンピナ，ジュデトゥル・プラホバ，ストラダ・ペトレ・リシウ・ヌマルル ４

(72)発明者 アレクサンドゥル，ノバク

ルーマニア国 ０４１３３４ ブカレスト，エスオーエス・オルテニティ・ヌマルル ２５４，ブロック １５１ エスセ １，アパルタメントゥル １８，セクター ４

(72)発明者 エレナ，ブルチャヌ

ルーマニア国 ０４０３６５ ブカレスト，ストラダ・コンスタンティン・ラドゥレス・モトゥル・ヌマルル ４，ブロック １ エスセ ２，アパルタメントゥル ５８，セクター ４