(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第6077472号
(P6077472)
(45)発行日　平成29年2月8日(2017.2.8)　(24)登録日　平成29年1月20日(2017.1.20)

(51)Int.Cl.		F I		
G06N 99/00	(2010.01)	G06N	99/00	１５３
G06F 21/62	(2013.01)	G06F	21/62	３１８
G06F 17/30	(2006.01)	G06F	17/30	２１０Ｄ

請求項の数 23　(全 25 頁)

(21)出願番号	特願2013-556831(P2013-556831)
(86)(22)出願日	平成24年2月29日(2012.2.29)
(65)公表番号	特表2014-511536(P2014-511536A)
(43)公表日	平成26年5月15日(2014.5.15)
(86)国際出願番号	PCT/US2012/027158
(87)国際公開番号	WO2012/118905
(87)国際公開日	平成24年9月7日(2012.9.7)
審査請求日	平成27年2月13日(2015.2.13)
(31)優先権主張番号	13/038,299
(32)優先日	平成23年3月1日(2011.3.1)
(33)優先権主張国	米国(US)

(73)特許権者　501113353
シマンテック　コーポレーション
Ｓｙｍａｎｔｅｃ　Ｃｏｒｐｏｒａｔｉｏｎ
アメリカ合衆国，カリフォルニア州　９４０４３，マウンテン　ビュー，エリス　ストリート　３５０
(74)代理人　100147485
弁理士　杉村　憲司
(74)代理人　100166213
弁理士　永久保　宅哉
(74)代理人　100139491
弁理士　河合　隆慶

最終頁に続く

(54)【発明の名称】機械学習を行うためのユーザインターフェース及びワークフロー

(57)【特許請求の範囲】
【請求項１】
機密データについての複数の陽性例及び機密データについての複数の陰性例を含むトレーニングデータセットを、ユーザインターフェースを介して、受信するステップと、
機械学習を用いて前記トレーニングデータセットを分析して、機械学習ベースド検出（machine learning-based detection (MLD)）プロファイルをトレーニングするステップであって、前記MLDプロファイルは新たなデータを機密データ又は非機密データとして分類するのに用いられる、ステップと、
前記ユーザインターフェースにおいて、前記MLDプロファイルについてのクオリティーメトリックを表示するステップと
前記クオリティーメトリックがクオリティー閾値を充足しなかった場合には：
ユーザ入力に応答して前記トレーニングデータセットを変更するステップと、
前記変更されたトレーニングデータセットを分析して前記MLDプロファイルを再トレーニングするステップと
を行うステップと
を備える方法。
【請求項２】
前記トレーニングデータセットを分析するステップは：
前記トレーニングデータセットについて特徴抽出を行って前記陽性例の特徴及び前記陰性例の特徴を備える特徴セットを生成するステップと、

前記トレーニングデータセットから分類モデルを生成するステップと、
前記クオリティーメトリックを算出するステップであって、前記クオリティーメトリックは偽陽性レーティング、偽陰性レーティング又はメモリ利用レーティングの少なくとも１つを含む、ステップと
を備える、請求項１に記載の方法。
【請求項３】
前記トレーニングデータセットを分析する前に前記ユーザインターフェースを介してメモリ割り当てに関してのユーザ選択を受信するステップであって、データのカテゴリゼーションについての前記メモリ利用レーティングは前記メモリ割り当てに準拠する、ステップ、をさらに備える請求項２に記載の方法。
【請求項４】
前記偽陽性レーティングが偽陽性閾値内である及び前記偽陰性レーティングが偽陰性閾値内である場合、展開操作を可能とするステップと、
前記展開操作を行うためのユーザ要求を前記ユーザインターフェースを介して受信するステップと、
前記ユーザ要求を受信したことに応答して、前記MLDプロファイルをDLPシステムのデータロスプリベンション（DLP）ポリシーに追加するステップと
をさらに備える請求項２に記載の方法。
【請求項５】
前記トレーニングデータセットから、偽陽性を起こしたデータ及び偽陰性を起こしたデータの少なくとも１つを、前記ユーザインターフェースにて特定するステップ、をさらに備える請求項２に記載の方法。
【請求項６】
機密データについての前記陽性例及び機密データについての前記陰性例に関してのカテゴリゼーション情報を受信するステップと、
前記トレーニングデータセットに追加すべきデータのカテゴリを前記ユーザインターフェースにて特定して前記クオリティーメトリックを向上させるステップと
をさらに備える請求項１に記載の方法。
【請求項７】
前記MLDプロファイルについての新たなクオリティーメトリックを前記ユーザインターフェースに表示するステップ
をさらに備える請求項１に記載の方法。
【請求項８】
前記トレーニングデータセットはDLPシステムのデータロスプリベンション（DLP）ポリシーにより収集されたものであり、機密データについての前記複数の陰性例は前記DLPポリシーにより機密ドキュメントとして誤分類されたドキュメントを含み、
前記MLDプロファイルを前記DLPポリシーへ展開するステップ
をさらに備える請求項１に記載の方法。
【請求項９】
前記MLDプロファイルのための感度閾値についての選択を、前記ユーザインターフェースを介して受信するステップと、
前記選択に基づいて前記MLDプロファイルのための感度閾値設定を制御するステップと
をさらに備える請求項１に記載の方法。
【請求項１０】
命令を含むコンピュータ可読媒体であって、前記命令が処理装置により実行されると、
機密データについての複数の陽性例及び機密データについての複数の陰性例を含むトレーニングデータセットを、ユーザインターフェースを介して、受信するステップと、
機械学習を用いて前記トレーニングデータセットを分析して、機械学習ベースド検出（MLD）プロファイルをトレーニングするステップであって、前記MLDプロファイルは新たなデータを機密データ又は非機密データとして分類するのに用いられる、ステップと、

前記ユーザインターフェースにおいて、前記MLDプロファイルについてのクオリティーメトリックを表示するステップと
前記クオリティーメトリックがクオリティー閾値を充足しなかった場合には：
ユーザ入力に応答して前記トレーニングデータセットを変更するステップと、
前記変更されたトレーニングデータセットを分析して前記MLDプロファイルを再トレーニングするステップと
を行うステップと
を備える方法を前記処理装置に行わせる、コンピュータ可読媒体。
【請求項１１】
前記トレーニングデータセットを分析するステップは：
前記トレーニングデータセットについて特徴抽出を行って前記陽性例の特徴及び前記陰性例の特徴を備える特徴セットを生成するステップと、
前記トレーニングデータセットから分類モデルを生成するステップと、
前記クオリティーメトリックを算出するステップであって、前記クオリティーメトリックは偽陽性レーティング、偽陰性レーティング又はメモリ利用レーティングの少なくとも１つを含む、ステップと
を備える、請求項１０に記載のコンピュータ可読媒体。
【請求項１２】
前記方法は：
前記トレーニングデータセットを分析する前に前記ユーザインターフェースを介してメモリ割り当てに関してのユーザ選択を受信するステップであって、データのカテゴリゼーションについての前記メモリ利用レーティングは前記メモリ割り当てに準拠する、ステップ
をさらに備える、請求項１１に記載のコンピュータ可読媒体。
【請求項１３】
前記方法は：
前記偽陽性レーティングが偽陽性閾値内である及び前記偽陰性レーティングが偽陰性閾値内である場合、展開操作を可能とするステップと、
前記展開操作を行うためのユーザ要求を前記ユーザインターフェースを介して受信するステップと、
前記ユーザ要求を受信したことに応答して、前記MLDプロファイルをDLPシステムのデータロスプリベンション（DLP）ポリシーに追加するステップと
をさらに備える、請求項１１に記載のコンピュータ可読媒体。
【請求項１４】
前記方法は：
前記トレーニングデータセットから、偽陽性を起こしたデータ及び偽陰性を起こしたデータの少なくとも１つを、前記ユーザインターフェースにて特定するステップ、
をさらに備える、請求項１１に記載のコンピュータ可読媒体。
【請求項１５】
前記方法は：
機密データについての前記陽性例及び機密データについての前記陰性例に関してのカテゴリゼーション情報を受信するステップと、
前記トレーニングデータセットに追加すべきデータのカテゴリを前記ユーザインターフェースにて特定して前記クオリティーメトリックを向上させるステップと
をさらに備える、請求項１０に記載のコンピュータ可読媒体。
【請求項１６】
前記方法は：
前記MLDプロファイルについての新たなクオリティーメトリックを前記ユーザインターフェースに表示するステップ
をさらに備える、請求項１０に記載のコンピュータ可読媒体。

【請求項１７】
前記トレーニングデータセットはDLPシステムのデータロスプリベンション（DLP）ポリシーにより収集されたものであり、機密データについての前記複数の陰性例は前記DLPポリシーにより機密ドキュメントとして誤分類されたドキュメントを含み、前記方法は：
前記MLDプロファイルを前記DLPポリシーへ展開するステップ
をさらに備える、請求項１０に記載のコンピュータ可読媒体。
【請求項１８】
前記方法は：
前記MLDプロファイルのための感度閾値についての選択を、前記ユーザインターフェースを介して受信するステップと、
前記選択に基づいて前記MLDプロファイルのための感度閾値設定を制御するステップと
をさらに備える、請求項１０に記載のコンピュータ可読媒体。
【請求項１９】
コンピューティング装置であって、
機械学習マネージャのための命令を格納するためのメモリと
前記命令を実行するための処理装置と
を備える、コンピューティング装置であって、
前記命令は前記処理装置に：
前記機械学習マネージャのためのユーザインターフェースを提供させ、
機密データについての複数の陽性例及び機密データについての複数の陰性例を含むトレーニングデータセットを、前記ユーザインターフェースを介して、受信させ、
機械学習を用いて前記トレーニングデータセットを分析させて、新たなデータを機密データ又は非機密データとして分類するのに用いられる機械学習ベースド検出（MLD）プロファイルをトレーニングさせ、
前記ユーザインターフェースにおいて、前記MLDプロファイルについてのクオリティーメトリックを表示させ、
前記クオリティーメトリックがクオリティー閾値を充足しなかった場合には：
ユーザ入力に応答して前記トレーニングデータセットを変更することと、
前記変更されたトレーニングデータセットを分析して前記MLDプロファイルを再トレーニングすることと
を行わせる
命令である、コンピューティング装置。
【請求項２０】
前記トレーニングデータセットを分析することは：
前記トレーニングデータセットについて特徴抽出を行って前記陽性例の特徴及び前記陰性例の特徴を備える特徴セットを生成することと、
前記トレーニングデータセットから分類モデルを生成することと、
前記クオリティーメトリックを算出することであって、前記クオリティーメトリックは偽陽性レーティング、偽陰性レーティング又はメモリ利用レーティングの少なくとも１つを含む、算出することと
を備える、請求項１９に記載のコンピューティング装置。
【請求項２１】
前記命令は、前記処理装置にさらに：
前記トレーニングデータセットから、偽陽性を起こしたデータ及び偽陰性を起こしたデータの少なくとも１つを、前記ユーザインターフェースにて特定させる、
請求項２０に記載のコンピューティング装置。
【請求項２２】
前記トレーニングデータセットはDLPシステムのデータロスプリベンション（DLP）ポリシーにより収集されたものであり、機密データについての前記複数の陰性例は前記DLPポリシーにより機密ドキュメントとして誤分類されたドキュメントを含み、

前記MLDプロファイルを前記DLPポリシーへ展開させる命令をさらに備える、請求項１９に記載のコンピューティング装置。
【請求項２３】
前記MLDプロファイルのための感度閾値についての選択を、前記ユーザインターフェースを介して受信させ、
前記選択に基づいて前記MLDプロファイルのための感度閾値設定を制御させる
命令をさらに備える請求項１９に記載のコンピューティング装置。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明の実施形態は、データロスプリベンションに関連し、より具体的には機械学習ベースド検出（MLD、machine learning-based detection）プロファイルをユーザが生成及び展開できるようにするインターフェースを提供するデータロスプリベンション（DLP）システムに関する。
【背景技術】
【０００２】
多くの組織では、機密データを特定するため及び機密データへのアクセスを制御するためにデータロスプリベンション（DLP）システムを施行している。典型的なDLPシステムはディープコンテンツ検査及び分析によって機密データを保護し、これには記述的テクノロジ及びフィンガープリンティングテクノロジが含まれる。記述的テクノロジはキーワード、表現又はパターン及びファイルタイプへのマッチを特定することによって並びに他のシグネチャベースド検出手法を行うことによって、機密データを保護する。フィンガープリンティングテクノロジはファイルの全体又は部分への完全一致を特定することによって機密データを保護する。組織の機密データの多くを保護することに関して効果的であれども、非構造化データ並びに製品のフォーミュラ、ソースコード及び営業レポート等の知的財産権を大量に取り扱う場合、フィンガープリンティングテクノロジ及び記述的テクノロジでは限界がある。
【０００３】
非構造化機密データをより正確に保護するために、一部のDLPシステムではベクトル機械学習（VML、vector machine learning）テクノロジの使用が検討されている。しかし、VMLは実装するのにとても複雑である。このため、VMLを用いる現行のDLPシステムでは、機械学習ベースド検出（MLD、machine learning-based detection）プロファイルをデザインするVMLについての専門家が顧客のために必要となっている。そして、顧客に渡されるDLPシステムは、顧客が変更できない既定のMLDプロファイルを有している。このようなDLPシステムでは、ユーザが自己のMLDプロファイルを生成するためのユーザインターフェース又はワークフローは何ら提供されない。
【発明の概要】
【０００４】
１つの実施形態では、機械学習（ML、machine learning）のためのユーザインターフェースを介して機密データについての陽性例及び機密データについての陰性例を含むデータのトレーニングセットをコンピューティング装置が受信する。コンピューティング装置はデータのトレーニングセットを機械学習を用いて分析して、新たなデータを機密データ又は非機密データとして分類（classify）するのに用いることができるMLDプロファイルをトレーニングする。コンピューティング装置はMLDプロファイルについてのクオリティーメトリックをユーザインターフェースに表示する。１つの実施形態では、MLDプロファイルには統計的データ分類モデル並びに陽性例の統計的に有意な特徴及び陰性例の統計的に有意な特徴を備える特徴セットが含まれ、クオリティーメトリックには偽陽性レーティング、偽陰性レーティング又はメモリ利用レーティングの少なくとも１つが含まれる。１つの実施形態では、コンピューティング装置は、データのトレーニングセットから、偽陽性を起こしたデータ及び偽陰性を起こしたデータの少なくとも１つをユーザインターフェー

スにて特定する。
【０００５】
１つの実施形態では、コンピューティング装置は、ドキュメントのトレーニングセットを分析する前に、ユーザインターフェースを介してメモリ割り当てに関してのユーザ選択を受信し、データのカテゴリゼーション（categorization）についてのメモリ利用レーティングはメモリ利用割り当てに準拠する。１つの実施形態では、コンピューティング装置は、偽陽性レーティングが偽陽性閾値内である及び偽陰性レーティングが偽陰性閾値内である場合、展開操作を可能とする。MLDプロファイルを展開することについてのユーザ要求を受信したことに応答して、コンピューティング装置はMLDプロファイルをDLPシステムのデータロスプリベンション（DLP）ポリシーに追加する。
【０００６】
１つの実施形態では、コンピューティング装置は、機密データについての陽性例及び機密データについての陰性例に関してのカテゴリゼーション情報を受信する。そして、コンピューティング装置は、データのトレーニングセットに追加すべきデータのカテゴリをユーザインターフェースにて特定してクオリティーメトリックを向上させることができる。１つの実施形態では、コンピューティング装置は、クオリティーメトリックがクオリティー閾値を充足しなかった場合、ユーザ入力に応答してデータのトレーニングセットを変更する。そして、コンピューティング装置は、変更されたデータのトレーニングセットを分析してMLDプロファイルを再トレーニングして、MLDプロファイルについての新たなクオリティーメトリックをユーザインターフェースに表示する。
【０００７】
１つの実施形態では、データのトレーニングセットは、DLPシステムのデータロスプリベンション（DLP）ポリシーにより収集されたものであり、機密データについての複数の陰性例はDLPポリシーにより機密ドキュメントとして誤分類されたドキュメントを含む。この実施形態では、コンピューティング装置は、その後MLDプロファイルをDLPポリシーに展開することができる。
【０００８】
１つの実施形態では、コンピュータ可読媒体が、プロセッサによる命令実行時に、機密データについての複数の陽性例及び機密データについての複数の陰性例を含むトレーニングデータセットを機械学習のためのユーザインターフェースを介して受信するように、該プロセッサを導く命令を含む。その後、プロセッサは、機械学習を用いてトレーニングデータセットを分析して、新たなデータを機密データ又は非機密データとして分類するのに用いることができる機械学習ベースド検出（MLD）プロファイルをトレーニングし、MLDプロファイルについてのクオリティーメトリックをユーザインターフェースに表示する。
【０００９】
１つの実施形態では、MLDプロファイルを生成するための方法が、機密データについての複数の陽性例及び機密データについての複数の陰性例を含むトレーニングデータセットを機械学習のためのユーザインターフェースを介して受信するステップと、機械学習を用いてトレーニングデータセットを分析して、新たなデータを機密データ又は非機密データとして分類するのに用いることができる機械学習ベースド検出（MLD）プロファイルをトレーニングするステップと、MLDプロファイルについてのクオリティーメトリックをユーザインターフェースに表示するステップとを備える。
【図面の簡単な説明】
【００１０】
後述の詳細な説明及び本発明の様々な実施形態についての添付の図面を参照することにより、本発明をより完全に理解することができる。
【図１】本発明の１つの実施形態による、例示的なシステムアーキテクチャの図である。
【図２】本発明の１つの実施形態による、データロスプリベンションエージェントのブロック図である。
【図３】本発明の１つの実施形態による、機械学習のブロック図である。

【図４】MLDプロファイルを生成及び展開するための方法についての１つの実施形態を示すフローチャートである。
【図５】本発明の実施形態による、MLDプロファイルを生成及び展開するためのユーザインターフェースについての様々な視点を示す図である。
【図６】本発明の実施形態による、MLDプロファイルを生成及び展開するためのユーザインターフェースについての様々な視点を示す図である。
【図７】本発明の実施形態による、MLDプロファイルを生成及び展開するためのユーザインターフェースについての様々な視点を示す図である。
【図８】本発明の実施形態による、MLDプロファイルを生成及び展開するためのユーザインターフェースについての様々な視点を示す図である。
【図９】本発明の１つの実施形態による、MLDプロファイル生成時におけるMLマネージャの様々な状態を示す状態図である。
【図１０】MLDプロファイルを生成して既存のDLPポリシーへMLDプロファイルを展開する方法についての１つの実施形態を示すフローチャートである。
【図１１】MLDプロファイルを含むDLPポリシーを用いてデータロスからコンピューティング装置を保護する方法についての１つの実施形態を示すフローチャートである。
【図１２】本明細書中の操作の１以上を行い得る例示的コンピュータシステムのブロック図である。
【発明を実施するための形態】
【００１１】
データロスプリベンション（DLP）システムのための機械学習ベースド検出（MLD）プロファイルを生成、展開及び管理するためのシステムと方法を説明する。本発明の実施形態において、システム及び方法は、ベクトル機械学習についての専門家でないユーザがMLDプロファイルを生成できるようにするユーザインターフェース及びワークフローを提供する。これによりDLPのためのMLDプロファイルを展開するためのコストが減少し、また、MLDプロファイルのコンフィグラビリティが向上する。さらに、これにより、DLPアドミニストレータが継続的にMLDプロファイルを改良できるようになる。
【００１２】
以下の説明では、様々な詳細を述べる。もっとも、本願の開示を得た当業者には、本発明がこれらの詳細を知らなくても実施できるものであると理解されるであろう。一部の場合においては、周知の構造及び装置は、本発明を不明瞭としないために、詳細を示さずにブロック図で示す。例えば、以下の説明ではエンドポイントDLPシステムにてMLDプロファイルを用いるための詳細を提供する。もっとも、当業者にとっては、本発明の実施形態はネットワークDLPシステム及びディスカバーDLPシステム（即ち記憶装置をスキャンして機密データを特定及び／又は分類する類いのDLPシステム）にも適用されることが明らかである。例えば、本発明の実施形態においてはエンタープライズネットワークの中を移動する機密データを検出するためのMLDプロファイルを生成することができる。
【００１３】
後述される詳細な説明の一部は、アルゴリズム及びコンピュータメモリ内のデータビットに対しての操作を表すシンボリック表記で提示される。これらのアルゴリズム的記述及び表記は、業務の内容をもっとも効果的に他の当業者に伝達するために、データ処理技術の当業者により用いられる手段である。ここでは及び一般的には、アルゴリズムとは、所望の結果へ至る自己一貫的なステップのシーケンスとして認識される。ステップとは、物理的な量に対して必要とされる物理的な操作のことである。必ずしもそうではないが、通常は、これらの量は格納、移転、合体、比較及び他の操作の対象とされ得る電気的又は磁気的信号の形をとる。主に慣例からして、これらの信号をビット、値、エレメント、シンボル、キャラクタ、ターム、数字等と呼ぶことが場合によっては便利である。
【００１４】
もっとも、これら全て及びこれらに類似の用語は適切な物理量と関連付けられるべきものであることに留意すべきであり、これらの量についての便利なラベルに過ぎない。以下

の説明から自明であるように、そうでないと具体的に宣言されない限り、本明細書中では、“受信する”“分析する”“表示する”“可能とする”“特定する”“変更する”等の用語を伴う説明は、物理量として表現されたコンピュータ装置のレジスタ及びメモリ内のデータを、操作及びコンピュータシステムのメモリ又はレジスタ若しくは情報を格納・伝送・表示する他の装置内の同様に表現されたデータに変換する、コンピュータシステム又は類似の電子的コンピューティング装置の動作又はプロセスを意味する。

【００１５】

本発明はこれらの操作を行うための装置にも関する。この装置は、必要とされる用途のために特に構築されるものであることができ、或いは、コンピュータ内に格納されたコンピュータプログラムにより選択的に起動又は再構成された汎用コンピュータであることができる。このようなコンピュータプログラムは、次のものには限定はされないが、フロッピーディスク、光学ディスク、CD-ROM及び光磁気ディスクを含む任意のタイプのディスク、リードオンリメモリ（ROM）、ランダムアクセスメモリ（RAM）、EPROM、EEPROM、磁気若しくは光学カード又は電子的命令を格納するのに適した他のあらゆるタイプの媒体等のコンピュータ可読記憶媒体に格納されることができる。

【００１６】

図１は、本発明の１つの実施形態による例示的システムアーキテクチャ１００を示す。システムアーキテクチャ１００は、エンドポイントサーバ１１５にネットワークされた複数のエンドポイント装置１０２A～１０２Cを含み、さらにエンフォースメントサーバ１２０にネットワークされている。

【００１７】

各エンドポイントサーバにネットワーク装置はパソコン（PC）、ラップトップ、携帯電話、タブレットコンピュータ又はユーザがアクセスできる他の任意のコンピューティング装置であることできる。各エンドポイント装置１０２A～１０２Cは複数の異なるデータロスベクトルを有する。各データロスベクトルは、エンドポイント装置からデータを移転できる経路である。データロスベクトルの例としては、光学ディスクにファイルを焼く行為、携帯可能ドライブ（例えば、携帯可能なユニバーサルシリアルバス（USB）ドライブ）にデータをコピーする行為、プリンタでデータを印字する行為、ファクシミリを通じてデータを送信する行為、電子メールの送信行為、インスタントメッセージの送信行為、画面コピー操作等がある。

【００１８】

エンドポイント装置１０２A～１０２Cは各々、エンドポイント装置のハードウェア及びソフトウェアを管理するオペレーティングシステム（OS）を実行している。OSは、例えば、Microsoft（登録商標）Windows（登録商標）、Linux（登録商標）、Symbian（登録商標）、Apple（登録商標）社のOS X（登録商標）、Solaris（登録商標）等であることができる。OS上では１以上のアプリケーションが実行されており、エンドポイント装置に含まれる、直接的に接続されている若しくはネットワークされているデータストアに存するデータに対してのアクセス、移動若しくは他の操作を伴う様々な操作を行っている。例えば、アプリケーションにはCD又はDVDバーニングアプリケーション、電子メールアプリケーション、ウェブブラウザ、インスタントメッセージアプリケーション、プリンティングアプリケーション、画面コピー機能が含まれ得る。１つの実施形態では、アプリケーションは、ユーザ命令を受信したことに応答して操作を行う。

【００１９】

各エンドポイント装置１０２A～１０２Cはデータストア１３５A～１３５Cに接続されていることができ、これはハードディスク、テープバックアップ、光学ドライブ、揮発性メモリ（例えば、ランダムアクセスメモリ（RAM））又は他の記憶装置であることができる。データストア１３５A～１３５Cは、エンドポイント装置１０２A～１０２Cとの関係で内蔵のもの又は外付けのものであることができる。１つの実施形態では、データストア１３５A～１３５Cはストレージエリアネットワーク（SAN）又はネットワークアタッチドストレージ（NAS）等のネットワークストレージに組み込まれることができる。１つの実施形

態では、データストア１３５A～１３５Cはリレーショナルデータベース等のデータベースに組み込まれることができる。データストア１３５A～１３５Cは機密情報を含むデータを含み得る。データは、ファイル（例えば、ドキュメント）、テーブル又は他のデータフォーマットを含むことができる。機密情報の例には、ソースコード、患者健康情報、保険請求、製品のフォーミュラ、法的書類、合併及び吸収に関する書類、営業レポート、社会保障番号、クレジットカード番号が含まれる。

【００２０】

各エンドポイント装置１０２A～１０２Cは、不正な目的によって機密（例えば、部外秘）情報がエンドポイント装置を離れないようにするために、データロスベクトルを監視するDLPエージェント１０６を含む。DLPエージェント１０６は、データロスベクトルをつうじて移動する際に及び／又はデータロスベクトルを通じてデータを送ることについての要求が受信された際にデータをスキャンすることができる。DLPエージェント１０６がデータロスベクトルを通じて移動するデータ又はデータロスベクトルを通じてデータを移動させることについての要求を検出した際には、DLPエージェント１０６はDLPポリシー１１０を実施してデータが機密データであるか（機密情報を含むか）を判断する。DLPポリシー１１０は、監視すべきコンテンツのタイプ（例えば、メッセージ、表示されたデータ、格納データ等）、どのように機密データを特定するか、及び／又は機密データを検出した際に行うべき動作を指定することができる。１つの実施形態では、DLPポリシー１１０はMLDプロファイル１１２を含む。DLPエージェント１０６は、MLDプロファイル１１２を用いてデータを処理する機械学習（ML）モジュール１０８を含む。MLDプロファイル１１２を用いてデータを処理することにより、MLモジュール１０８はデータが機密データかを判断する。

【００２１】

一部のタイプのDLP検出手法については、DLPエージェント１０６はデータをエンドポイントサーバ１１５へ送り、エンドポイントサーバ１１５に含まれるグローバルDLP検出エンジン１２２がデータに機密情報が含まれるかを判断する。一旦グローバルDLP検出エンジン１２２がファイル又は他のデータが機密情報を含むものであると判断すると、エンドポイントサーバ１１５はDLPエージェント１０６にデータが機密データであるか否かを宣言するメッセージを送り返す。そして、データがコンフィデンシャルな情報を含む場合にDLPエージェント１０６は、DLPポリシー１１０をエンフォースするための１以上の動作を行うことができる。１つの実施形態では、グローバルDLP検出エンジン１２２は、MLモジュール１０８及びMLDプロファイル１２８を含むDLPポリシー１２６を含む。DLPポリシー１２８及びMLDプロファイル１２８は、DLPポリシー１１０及びMLDプロファイル１１２とは異なることができる。

【００２２】

１つの実施形態では、エンドポイントサーバ１１５はDLPポリシー違反に関するデータのアグレゲータ（例えば、インシデントレポートのアグレゲータ）として機能する。エンドポイントサーバ１１５は、各エンドポイント装置からそのようなデータを収集して、収集されたデータを、分析のためにエンフォースメントサーバ１２０に報告することができる。

【００２３】

エンフォースメントサーバ１２０はDLPポリシーを管理する。これは、（例えば、アドミニストレータの入力に基づいて）DLPポリシーを生成すること及び変更することを含む。そして、エンフォースメントサーバ１２０はDLPポリシーをエンドポイントサーバ１１５及び／又はエンドポイント装置１０２へと伝播させることができる。また、エンフォースメントサーバ１２０はDLPレスポンスルールを生成してこれをエンドポイントサーバ１１５及び／又はエンドポイント装置１０２へと伝播させることもできる。DLPレスポンスルールは、DLPポリシー違反の際に、エンドポイント装置１０２及び／又はエンドポイントサーバ１１５がとるべき動作を指定する。エンドポイント装置が取り得る動作の例には、アドミニストレータに通知を送ること、データロスベクトルを通じデータがエンドポイ

ント装置１０２A～１０２Cから離脱することを防止すること、あらゆるデータロスベクトルを通じてエンドポイント装置からデータを移動させないようにするためにエンドポイント装置をロックダウンすること、エンドポイント装置からデータが移動される際にデータを暗号化すること等が含まれる。

【００２４】

１つの実施形態では、エンフォースメントサーバ１２０は機械学習（ML）マネージャ１３０を含む。MLマネージャ１３０は、ユーザがMLDプロファイルを生成及び展開するためのユーザインターフェース及びワークフローを提供する。MLマネージャ１３０については、図３を参照してより詳しく後述する。

【００２５】

図２は、本発明の１つの実施形態によるデータロスプリベンションエージェント２０５のブロック図である。DLPエージェント２０５は、異なるデータロスベクトル、アプリケーション、データ等を監視して、エンドポイント装置からデータを移動させようとする操作を検出することができる。ユーザにより開始される操作には、例えば、エンドポイント装置の任意の記憶装置上の制限付きデータベースのデータについてセーブ又はアクセスを行うこと、制限付きデータベースのデータをアプリケーション内で使用すること、コンフィデンシャルなデータをプリントすること、コンフィデンシャルなデータをネットワーク通信プロトコルで使用すること等が含まれ得る。

【００２６】

DLPエージェント２０５は、１以上のポリシー違反ディテクタを含むことができ、各々のそれは異なるDLPポリシー２５０及び／又はDLPポリシー２５０内の異なるプロファイル２５５、２６０、２６５を処理して、機密データを特定及び／又は保全することができる。DLPポリシー２５０は、高まったデータロスリスクを表す基準を含むことができる。DLPポリシー２５０は、DLPポリシー２５０内に含まれた基準の１以上が充足された場合に、違反されたことになる。基準の例には、ユーザ状態（例えば、ユーザがそのファイルに対してのアクセス権を有するか）、ファイルロケーション（例えば、コピーされようとしているファイルがコンフィデンシャルなデータベースに格納されているか）、ファイルコンテンツ（例えば、ファイルが機密情報を含むか）、時間（操作が通常の営業時間内に要求されているか）、データロスベクトル、操作を試みているアプリケーション等が含まれる。

【００２７】

DLPポリシー２５０は、１以上のプロファイル２５５、２６０、２６５を含むことができる。各プロファイルは、機密データを特定するのに用いることができる。１つの実施形態では、DLPポリシー２５０は記述的コンテンツマッチング（DCM）プロファイル２５５を含む。DCMプロファイル２５５は、サーチされるべき１以上のキーワード及び／又は正規表現を定義する。例えば、DCMプロファイル２５５は、正規表現を用いて社会保障番号を定義することができる。DCMプロファイル２５５を用いて、DLPエージェント２０５は、スキャンされたデータに含まれる何らかの情報がキーワード及び／又は正規表現にマッチするかを判断する。マッチが発見された場合、データが機密情報を含むものと判断されることができる。

【００２８】

１つの実施形態では、DLPポリシー２５０はイグザクトデータマッチング（EDM）プロファイル及び／又はインデクスドドキュメントマッチング（IDM）プロファイル２６０を含む。イグザクトデータマッチング（EDM）は、データベースレコード等の典型的に構造化されたフォーマットをとるデータを保護するのに用いることができる。インデクスドドキュメントマッチング（IDM）は、Microsoft（登録商標）Word若しくはPowerPoint（登録商標）ドキュメント又はCADドローイング等の非構造化データを保護するのに用いることができる。EDM及びIDMの両方では、データを保護することを望んでいる組織によって機密データがまず特定され、及びその後、進行形での正確な検出に資するためにフィンガープリンティングが行われる。１つの実施形態では、フィンガープリンティング処理は、テキス

トデータをアクセス及び抽出すること、それを正規化すること、及び不可逆ハッシュを用いてそれをセキュアすること、を含む。ファイル又は他のデータをスキャンすべき場合、そのファイル又はコンテンツについてのフィンガープリント（例えば、ハッシュ）が生成され、格納されているフィンガープリントと比較される。マッチが発見された場合には、スキャンされたファイルは機密データを含むものとして特定される。

【００２９】

１つの実施形態では、DLPポリシー２５０は機械学習ベースド検出（MLD）プロファイル２６５を含む。ベクトル機械学習及び他のタイプの機械学習を用いて、Microsoft（登録商標）Word、PowerPoint（登録商標）やCADドローイング等の非構造化データを保護することができる。MLDプロファイル２６５は、トレーニングデータセット２７０、分類モデル２７５及び特徴セット２８０を含むことができる。トレーニングデータセット２７０は、機密データについての陽性例及び機密データについての陰性例の集合である。トレーニングデータセット２７０は、MLマネージャによって処理されて分類モデル２７５及び特徴セット２８０が生成される。分類モデル２７５は、データ分類のための統計的モデルであり、境界特徴を表すサポートベクトルのマップを含む。特徴セット２８０は、リスト等のデータ構造であり、トレーニングデータセット２７０から抽出された複数の特徴を含む。１つの実施形態では、各特徴はトレーニングデータセット２７０からのデータに含まれているワードである。

【００３０】

ポリシー違反ディテクタの１例は、機械学習モジュール２２５である。MLモジュール２２５は、MLDプロファイル２６５及び未分類データ（例えば、ファイル２３５）を入力としてとりデータについての分類を出力するMLエンジン２３０を含む。MLエンジン２３０は、分類モデル２７５及び特徴セット２８０を用いて入力データを処理する。したがって、MLモジュール２２５は機密データと非機密データとを区別するのにMLDプロファイル２６５を用いることができる。

【００３１】

ポリシー違反レスポンダ２２０は、DLPポリシー違反が検出された場合、１以上のDLPレスポンスルール２４５を適用する。各DLPレスポンスルール２４５は、１以上のDLPポリシー２５０と関連付けられることができる。各DLPレスポンスルール２４５は、関連付けられたDLPポリシー２５０の違反に応答してポリシー違反レスポンダ２２０がとるべき１以上の動作を含む。一旦DLPポリシー２５０の違反が発見されると、ポリシー違反レスポンダ２２０が、どのDLPレスポンスルールが違反されたDLPポリシー２５０に関連付けられているかを判断することができる。その後、レスポンスルール２４５に含まれる１以上の動作が行われることができる。行われる動作の例には、アドミニストレータに通知を送ること、データロスベクトルを通じてデータがエンドポイント装置から離脱することを防止すること、あらゆるデータロスベクトルを通じてエンドポイント装置からデータを移動させないようにするためにコンピュータをロックダウンすること、エンドポイント装置からデータが移動される際にデータを暗号化すること等が含まれる。

【００３２】

インシデントレポートジェネレータ２１５は、違反されたDLPポリシー２５０及び違反されたDLPポリシー２５０に関連する事情を記録するインシデントレポート２４０を生成することができる。インシデントレポートジェネレータ２１５はエンドポイント装置で生じた及び／又は特定のユーザによって試みられたポリシー違反の一部又は全部についてのインシデントレポート２４０の記録を維持する。ユーザは、例えばユーザログインに基づいて特定されることができる。違反されたDLPポリシーを特定するのに加えて、各インシデントレポート２４０は、ポリシー違反についての状況を示すこともできる。例えば、インシデントレポート２４０は、ポリシー違反と関連付けられるアプリケーション、ユーザ、データロスベクトル、機密データのタイプ（例えば、社会保障番号、クレジットカード番号等）等を特定することができる。インシデントレポートジェネレータ２１５は、いつポリシー違反が起きたかを示すタイムスタンプを含めることもできる。

【００３３】
図３は、本発明の１つの実施形態による、機械学習（ML）マネージャ３０５のブロック図である。MLマネージャ３０５は、MLDプロファイルトレーナ３２５、MLDプロファイルテスタ３２０及び／又はMLDプロファイルデプロイヤ３１５を含む。１つの実施形態では、MLマネージャ３０５はユーザインターフェース３１０も含む。代替的な実施形態では、１以上のMLDプロファイルトレーナ３２５、MLDプロファイルテスタ３２０又はMLRプロファイルトレーナ３２５を１つのモジュールに組み合わせるか複数のモジュールに分割することができる。
【００３４】
MLDプロファイルトレーナ３２５は、トレーニングデータセット３５２に基づいてMLDプロファイル３６５をトレーニングする。MLDプロファイルトレーニングとは、トレーニングデータセットからコンテンツを抽出して、コンテンツに対して統計的分析を行って、分類モデル及び特徴セットを生成するプロセスをいい、これら双方について詳しく後に述べる。ユーザ（例えば、DLPアドミニストレータ）がトレーニングデータセットにおいて用いるべきデータを指定することができる。１つの実施形態では、ユーザが機密データについての陽性例（陽性データ３４５）及び機密データについての陰性例（陰性データ３５０）を選択して、トレーニングデータセット３５２へこれらを加える。これは、ユーザインターフェース３１０を介して行われることができる。代替的には、ユーザは、標準的なファイルシステムインターフェース（例えば、Microsoft（登録商標）Explorer（登録商標））を介して、ファイルを陽性データフォルダ及び陰性データフォルダに追加できる。データは、トレーニングデータセットに、個々のファイル（例えば、ドキュメント）として又は単一の圧縮ファイル（例えば、zipファイル）のコンポーネントとして、追加されることができる。
【００３５】
１つの実施形態では、トレーニングデータセット３５２のためのデータは、インシデントレポート３６０から抽出される。インシデントレポート３６０は、DLPポリシー３８５のエンフォースメント中に、既存のDLPポリシーについて生成されたものかもしれない。インシデントレポート３６０は、機密データについて操作が行われた又はその操作を行うことが要求されたときの事情を特定することができる。インシデントレポートは、機密データの真性該当事例を含むことができ、また、非機密データが機密データとして分類された偽陽性を含むこともできる。インシデントレポートと関連付けられる又は関連付けられない、他の履歴データも、トレーニングデータセットとして用いられることができる。履歴データは、機密データの真性該当事例、偽陽性、非機密データの真性該当事例、及び／又は偽陰性を含むことができる。
【００３６】
１つの実施形態では、MLDプロファイルトレーナ３２５は、既存のMLDプロファイルについてインクリメンタル型のトレーニングを行う。インクリメンタル型のトレーニングでは、MLDプロファイルトレーナ３２５は、MLDプロファイルが最後にトレーニングされてから後に生成されたインシデントレポートに基づいた、新たな陽性データ及び／又は陰性データを、トレーニングデータセットに追加する。MLDプロファイルトレーナ３２５は、自動的に又はユーザ入力に応答して、インクリメンタル型のトレーニングを行うことができる。１つの実施形態では、既定のスケジュールに従ってインクリメンタル型のトレーニングが行われる。例えば、MLDプロファイルトレーナ３２５は、MLDプロファイルについて、毎日、毎週、毎月等のように、定期的にトレーニングを行うことができる。
【００３７】
１つの実施形態では、MLDプロファイルトレーナ３２５は、トレーニングデータセットに閾値に該当する数のドキュメントが追加されるまで、トレーニングデータセット３５２についてのMLDプロファイル３２５を生成しない。１つの実施形態では、陽性データ３４５についての閾値及び陰性データ３５０についての閾値を追加するものとする。閾値は、例えば、５０件の陽性ドキュメント及び５０件の陰性ドキュメントとすることができる。

１つの実施形態では、MLマネージャ３０５によって、最大ドキュメントサイズ（例えば、１５MB、３０MB等）が強制される。最大ドキュメントサイズより大きい如何なるドキュメントもトレーニングデータとして使用することについて棄却することができる。最大ドキュメントサイズをユーザによって選択可能とすることができる。

【００３８】

１つの実施形態では、MLDプロファイルトレーナ３２５は、モデルジェネレータ３３０、特徴エクストラクタ３３５及びクオリティーアナライザ３４０を含む。特徴エクストラクタ３３５は、トレーニングデータセット３５２内の機密データについての陽性例及び機密データについての陰性例分析を行い、陽性データ及び陰性データ中での特徴（例えば、ワード）の出現頻度を決定する。その後、特徴エクストラクタ３３５は、例えば出現頻度に基づいて、陽性特徴及び陰性特徴をランク付けする。１つの実施形態では、特徴エクストラクタ３３５は、“the”、“it”、“and”等のありきたりのワードをフィルターアウトする。特徴エクストラクタ３３５はその後、特徴セット３７５のために、もっとも高くランクされた特徴を選択する。

【００３９】

１つの実施形態では、特徴エクストラクタ３３５は、中国キャラクタ（漢字）等のキャラクタベースドアルファベットについては、キャラクタから特徴を生成する。特徴エクストラクタ３３５は、各キャラクタについて特徴を生成し、また、隣接するキャラクタの組についても特徴を追加的に作成する。例えば、複数のキャラクタ Ξ λ については、特徴エクストラクタ３３５はΞとλとΞλとについて特徴を生成する。

【００４０】

特徴セット３７５に追加される特徴の数はメモリ割り当てに基づくことができ、これはMLDプロファイルトレーナ３２５により自動的に選択され又はユーザにより選択されることができる。メモリ割り当てが増大するにつれ、特徴セット３７５に含まれる特徴の数も増大し、MLDプロファイルの正確性を向上させ得る。メモリ割り当ては、例えば、およそ３０MBからおよそ１００MBの間で可変とすることができる。１つの実施形態では、メモリ割り当ては、ハイ、ミディアム又はローとして選択可能である。代替的には、具体的なメモリ割り当てを選択することができる（例えば、４３MB）。結果的なMLDプロファイル３６５のサイズは、トレーニングドキュメントの数及びメモリ割り当て設定に比例する。１つの実施形態では、DLPエージェントにより実施されるMLDプロファイル３６５にはより低いメモリ割り当てが用いられ、グローバルDLP検出エンジンにより実施されるMLDプロファイル３６５にはより高いメモリ割り当てが用いられる。

【００４１】

１つの実施形態では、特徴エクストラクタ３３５は、特徴セット３７５を選択するのに、ターム頻度―ドキュメント逆頻度（TF-IDF、term frequency-inverse document frequency）アルゴリズムを用いる。代替的には、特徴エクストラクタ３３５は、segment-set term frequency-inverse segment-set frequency (STF-ISSF)やsegment-set term frequency-inverse document frequency (STF-IDF)等の他の特徴抽出アルゴリズムを用いることができる。１つの実施形態では、特徴エクストラクタ３３５が用いる特徴選択アルゴリズムは、ユーザにより選択可能とされる。また、特徴エクストラクタ３３５は、複数回特徴抽出を行い、各回において異なる特徴抽出アルゴリズムを用いることができる。異なるアルゴリズムを用いて生成された特徴セットは、各々異なる分類モデルを生成するのに用いることができ、クオリティーアナライザ３４０によりテストされることができる。最良のクオリティーメトリックを有する特徴セットを保存して他を破棄することができる。

【００４２】

特徴エクストラクタ３３５が特徴セット３７５を生成した後は、モデルジェネレータ３３０が、特徴セット３７５及びトレーニングデータセット３５２に基づいて、分類モデル３８０を生成する。分類モデル３８０は、境界特徴を表すサポートベクトルのマップを含む、データ分類のための統計的モデルである。境界特徴は特徴セット３７５から選択することができ、特徴セット３７５で最も高くランクされた特徴を表すことができる。

【００４３】
一旦特徴エクストラクタ３３５が特徴セット３７５を生成して、モデルジェネレータ３３０が分類モデル３８０を生成すると、MLDプロファイル３６５は完成する。MLDプロファイル３６５は、特徴セット３７５、分類モデル３８０及び／又はトレーニングデータセット３７０を含むことができる。MLDプロファイル３６５は、ユーザ定義の設定を含むこともできる。１つの実施形態では、ユーザ定義の設定は、感度閾値（信頼水準閾値ともいう）を含む。感度閾値は、例えば、７５％、９０％等と設定できる。MLエンジンが、ドキュメントを機密又は機密でないと分類するのにMLDプロファイル３６５を用いる場合、MLエンジンは分類に信頼値を付与することができる。ドキュメントについての信頼値が１００％の場合、ドキュメントが機密（又は機密でない）との判断は、例えば、信頼値が５０％の場合に比してより確実なものである。信頼値が感度閾値よりも少ない場合、ドキュメントが機密ドキュメントと分類されども、インシデントが生成されないようにできる。この機能により、偽陽性及び／又は偽陰性を更に制御又は削減することについて、ユーザを支援できる。MLエンジンが、トレーニングで見られたことのないタイプのドキュメントを分類しようとしている場合、ドキュメントが陽性及び／又は陰性であることについてはとても低い信頼を持つことになる。このような場合においては、偽陽性の頻度を減少させるために感度閾値を用いることができる。１つの実施形態では、MLDプロファイルトレーナ３２５は、トレーニングに基づいて、自動的にMLDプロファイル３６５のための感度閾値を選択する。

【００４４】
１つの実施形態では、クオリティーアナライザ３４０はMLDプロファイル３６５のクオリティーを分析して、MLDプロファイル３６５についての１以上のクオリティーメトリックを生成する。クオリティーメトリックには、偽陽性レーティング（MLDプロファイル３６５によって機密データとして誤分類された機密データについての陰性例）、偽陰性レーティング（MLDプロファイル３６５によって非機密データとして誤分類された機密データについての陽性例）、及び／又はメモリ利用レーティング（MLDプロファイル３６５によって利用されるメモリ量）を含めることができる。クオリティーアナライザ３４０は、クオリティーメトリックを１以上のクオリティー閾値と比較することができる。これらには、偽陽性閾値、偽陰性閾値、及び／又はメモリ利用閾値が含まれ得る。１つの実施形態では、偽陽性閾値は５％とされ、また、偽陰性閾値が５％とされる。代替的には、他の偽陽性及び／又は偽陰性閾値を用いることができる。偽陽性レーティングが偽陽性閾値を上回る場合、偽陰性レーティングが偽陰性閾値を上回る場合、又はメモリ利用レーティングがメモリ利用閾値を上回る場合、MLDプロファイル３６５は展開されるに相応しくないかもしれない。１以上のクオリティー閾値をMLDプロファイル３６５が超えていない場合、MLマネージャ３０５はMLDプロファイル３６５の展開を許可しないことができる。

【００４５】
トレーニングデータセット３５２を変更して、及び、特徴セット３７５及び分類モデル３８０を再算出することによって、MLDプロファイル３６５を変更することができる。新たな陽性データ３４５を追加すること、新たな陰性データ３５０を追加すること、陽性データ３４５のインスタンスを削除すること、及び／又は陰性データ３５０のインスタンスを削除することによって、トレーニングデータセット３５２を変更することができる。１つの実施形態では、クオリティーアナライザ３４０が、偽陽性を起こした陰性データ３５０から、具体的なファイルやドキュメント等を特定し、また、偽陰性を起こした陽性データ３４５から、具体的なファイルやドキュメント等を特定する。ユーザはこの情報を検討して、トレーニングデータセットに追加すべき追加的データを判断することができる。トレーニングデータセット３５２において、特定のカテゴリのドキュメントが少なすぎた場合があり得る。例えば、ユーザはソースコードを保護することを望む一方、製品ドキュメンテーションがMLDプロファイル３６５によりソースコードとしてクロス分類されたかもしれない。ユーザは、陰性データセットに製品ドキュメンテーションの追加的な例を追加することによって、これを是正できる。機密又は非機密と認識・分類され得るデータのカ

テゴリの例は、ソースコード、レシピ、法的文書、製品ドキュメンテーション、医療履歴文書、保険文書、製品フォーミュラ、患者健康情報等を含む。

【００４６】

１つの実施形態では、ユーザは、ユーザがトレーニングデータセットに追加する各ファイル（例えば、ドキュメント）について特定のカテゴリを指定することができる。そして、クオリティーアナライザ３４０は、最も多くの偽陽性及び／又は最も多くの偽陰性を起こしたドキュメントカテゴリを特定することができる。１つの実施形態では、クオリティーアナライザ３４０は、MLDプロファイル３６５のクオリティーを向上させるためにユーザが追加すべきドキュメントの特定のカテゴリを提案する。

【００４７】

１つの実施形態では、MLマネージャ３０５は、以前に生成されたMLDプロファイルに加えられた変更を含む、チェンジレポートを維持する。チェンジレポートは、以前に生成されたMLDプロファイルと最近変更されたMLDプロファイルのクオリティーメトリックにおける差を含むこともできる。チェンジレポートは、ユーザが変更をアクセプトするために又は変更をロールバックして以前のMLDプロファイルに戻るために、ユーザに表示することができる。

【００４８】

一旦MLDプロファイル３６５が展開に相応しいものとなると（例えば、クオリティーメトリックがクオリティー閾値内のものとなった場合）、MLDプロファイルデプロイヤ３１５はMLDプロファイル３１５を展開する。１つの実施形態では、MLDプロファイルデプロイヤ３１５は、該MLDプロファイルを既存のDLPポリシー３８５に追加する。代替的には、MLDプロファイルデプロイヤ３１５は、新たなDLPポリシーを生成して、新たなDLPポリシー３８５にMLDプロファイル３６５を追加することができる。

【００４９】

１つの実施形態では、MLマネージャ３０５はVMLテスタ３２０を含む。VMLテスタは、追加的データをもってMLDプロファイルをテストする。１つの実施形態では、MLDプロファイルテスタは、既定のテスティングデータのセットについて陰性テスティングを行う。既定のテスティングデータのセットは、機密情報を含まないものとして知られるデータを大量に（例えば、10,000個のドキュメント）含むことができる。MLDプロファイルテスタ３２０は、追加的な陽性データ及び／又は陰性データを含み得る、ユーザ選択データについてMLDプロファイルをテストすることもできる。

【００５０】

図４は、MLDプロファイルを生成及び展開する方法４００についての１つの実施形態を図示するフローチャートである。方法４００は、ハードウェア（回路、専用のロジック等）、ソフトウェア（汎用コンピュータシステム又は専用機で実行されるもの）又は両者の組合せを含むことができる処理ロジックにより行われる。方法４００は、図１のエンフォースメントサーバ１２０上で実行されているMLマネージャ１３０のようなMLマネージャによって行われることができる。方法４００は下記においてMLマネージャにより行われるものと説明されるが、方法４００は他の処理ロジックによっても行われることができる。

【００５１】

図４を参照するに、ブロック４０５では、MLマネージャが、新たなMLDプロファイル又は変更されるべき既存のMLDプロファイルのための一時的ワークスペースを、生成する。１つの実施形態では、IDM又はEDMプロファイル等の他のプロファイルが既に実行されているような、空のMLDプロファイルが、既存のDLPポリシーについて生成される。他の実施形態では、まだ展開されていない新たなDLPポリシーについて、空のMLDプロファイルが生成される。代替的には、一時的ワークスペースにて既存のMLDプロファイルがオープンされる。１つの実施形態では、機械学習についてのインターフェースを介して、新たなMLDプロファイルを作成せよとのユーザ要求又は既存のMLDプロファイルを変更せよとのユーザ要求、に応答して一時的ワークスペースが生成される。１つの実施形態では、新たなMLDプロファイルは、特定のカテゴリのデータを保護するためのものである。例えば、MLDプ

ロファイルはソースコードを保護するため、患者情報を保護するため、販売データを保護するためのもの等とすることができる。

【００５２】

図５は、空の一時的ワークスペースを示す、本発明の１つの実施形態による、機械学習のためのユーザインターフェースの第１の表示５００を示す。示されているように、一時的ワークスペースは、陽性ドキュメントをアップロードするための陽性ボタン５０５及び陰性ドキュメントをアップロードするための陰性ボタン５１０を有する。ユーザが陽性ボタン５０５又は陰性ボタン５１０を選択することに応答して、MLマネージャはファイルブラウザウィンドウを開くことができる。そして、ユーザはファイルブラウザウィンドウをナビゲートして、アップロードのためのドキュメントを選択することができる。

【００５３】

１つの実施形態では、ユーザインターフェースはメモリ割り当てボタン５１５を含む。ユーザがメモリ割り当てボタン５１５を選択することに応答して、MLマネージャはメモリ割り当てについての選択肢をユーザに提示するウインドウを開く。１つの実施形態では、ユーザは、ハイ、ミディアム及びローのメモリ割り当てを選択することができる。各メモリ割り当ては、特定のメモリ利用閾値と関連付けられることができる。代替的には、ユーザは、具体的なメモリ割り当て（例えば、１２MB、５４MB等）を選択することができる。１つの実施形態では、ユーザインターフェースは、押すとプロファイル名及び／又はプロファイル説明をタイプインできるウインドウを開く追加のボタン５１５を、含む。

【００５４】

図４に戻るに、方法４００のブロック４１０では、MLマネージャはトレーニングデータセットを受信する。１つの実施形態では、ユーザが、トレーニングデータセットのためのデータを、ユーザインターフェースを介して、選択する。トレーニングデータセットは、機密データについての陽性例及び機密データについての陰性例の両方を含む。トレーニングデータセットは、複数のドキュメントを含むことができる。１つの実施形態では、ユーザが、各ドキュメントについてのカテゴリ（例えば、ソースコード、販売データ、医療記録等）を指定する。１つの実施形態では、MLマネージャが、トレーニングデータセット内の各ドキュメントについてドキュメントサイズをチェックする。MLマネージャは、最大ドキュメントサイズを超えるMLドキュメントを棄却することができる

【００５５】

図６は、トレーニングデータセットを示す、本発明の１つの実施形態による、図５のユーザインターフェースについての第２の表示６００を示す。トレーニングデータセットの各ドキュメントは、ドキュメントの複数の属性と共に表示されることができる。１つの実施形態では、表示されるドキュメント属性は、ドキュメントタイプ６０５（即ちドキュメントが陽性ドキュメントであるか陰性ドキュメントであるか）、ドキュメントの名前６１０、ドキュメントの日付６１５（即ち、ドキュメントがアップロードされた日付）、及びドキュメントの作成者６２０を含む。各ドキュメントには削除ボタン６２５が付されることもできる。削除ボタン６２５を選択することにより、ユーザは、特定のドキュメントをトレーニングデータセットから除くことができる。１つの実施形態では、ドキュメントカテゴリも示される。ユーザは、各ドキュメントに、ドキュメントカテゴリを付与することができる。１つの実施形態では、一時的ワークスペースのタブにある×印ボタンをクリックすることにより、プロファイルに加えられた変更の全てがロールバックされる。そして、プロファイルは、最後にあった機能的状態に留まり続ける。

【００５６】

閾値に届く数の陽性ドキュメント及び陰性ドキュメントがトレーニングデータセットに追加されると（例えば、各タイプについて２０個のドキュメント、や各タイプについて５０個のドキュメント等）、プロファイルをトレーニングする操作が利用可能となる。１つの実施形態では、閾値に届く数の陽性ドキュメント及び陰性ドキュメントが追加されると、“プロファイルをトレーニングする”ボタン６３０がアクティブになる。ユーザは、MLDプロファイルをトレーニングするために（例えば、MLDプロファイルのために特徴セット

及び分類モデルを生成するために）、“プロファイルをトレーニングする”ボタン６３０を選択することができる。

【００５７】

図４に戻るに、方法４００のブロック４１５では、MLマネージャは、メモリ割り当てについての選択を受信する。ブロック４２０では、MLマネージャは、機械学習（例えば、ベクトル機械学習）を用いてトレーニングデータセットを分析してMLDプロファイルをトレーニングする。１つの実施形態では、MLマネージャは、トレーニング中においては、MLDプロファイルに対して書き込みロックを行う。１つの実施形態では、MLDプロファイルのトレーニングには、特徴抽出を行うこと（ブロック４２１）、分類モデルを生成すること（ブロック４２２）及び分類モデル及び特徴セットのクオリティーを判断すること（ブロック４２３）が含まれる。ブロック４２５では、MLマネージャは分析の結果をユーザインターフェースに表示する。結果には、偽陽性レーティング、偽陰性レーティング、メモリ利用レーティング、抽出に失敗した陽性ドキュメント、及び抽出に失敗した陰性ドキュメント等の１以上のクオリティーメトリックが含まれることができる。１つの実施形態では、ユーザは、失敗した抽出情報をクリックしてどのドキュメントについて失敗があったかを知ることができる。

【００５８】

図７は、トレーニングされているMLDプロファイルを示す、図５のユーザインターフェースについての第３の表示７００を図示する。１つの実施形態では、ユーザインターフェースが、特徴抽出、正確性算定、モデル作成、及び最終処理を含む、MLDプロファイルトレーニングの各ステップを表示する。MLマネージャがMLDプロファイル生成においてどの段階にあるかを示すためにMLDプロファイルトレーニングの現在ステップをハイライトすることができる、１つの実施形態では、プロファイルがトレーニングされている間、一時的ワークスペースはロックアップされる。また、プロファイルがトレーニングされている間、メモリ割り当てを調整することはできない。これにより、正確なトレーニング結果が得られることが保証される。１つの実施形態では、ユーザは、トレーニングをキャンセルする選択肢をいつでも選択することによって、トレーニングを停止することができる。

【００５９】

図８は、MLDプロファイルトレーニング結果を示す、図５のユーザインターフェースについての第４の表示８００を図示する。１つの実施形態では、MLDプロファイルトレーニング結果には、陽性ドキュメントカウント８０５、陰性ドキュメントカウント８１０及び総ドキュメントカウント８１５が含まれる。１つの実施形態では、MLマネージャが特徴セットに含まれる特徴（例えば、ワード等）のリストを表示する。トレーニング結果には、MLDプロファイルについての１以上のクオリティーメトリックも含まれる。１つの実施形態では、クオリティーメトリックには、偽陽性レーティング８２０、偽陰性レーティング８２５及びメモリ利用レーティング８３０が含まれる。ユーザは偽陽性レーティング８２０を選択して、偽陽性を起こした具体的なドキュメント等の偽陽性についての追加的情報を閲覧することができる。また、ユーザは、偽陰性レーティング８２５を選択して偽陰性を起こした具体的なドキュメント等の偽陰性についての追加的情報を閲覧することができる。クオリティーメトリックがクオリティー閾値内にある場合、“トレーニングを展開する”ボタン８３５がアクティブになることができる。ユーザは、“トレーニングを展開する”ボタン８３５を選択してMLDプロファイルを展開することができる。ユーザは、“トレーニングを棄却”ボタン８４０を選択してMLDプロファイルを棄却することができる。

【００６０】

図４に戻るに、ブロック４３０では、分析結果からしてMLDプロファイルが１以上の展開条件を充足しているといえるかを、MLマネージャが判断する。結果が展開条件を満たす場合、方法は、ブロック４３５へと進む。そうでなければ、方法はブロック４４０へと進む。

【００６１】

ブロック４３５では、MLマネージャがMLDプロファイルの展開操作を可能とする。ブロ

ック４５０では、MLマネージャは（例えば、ユーザがユーザインターフェースの展開ボタンを押下することに基づく）展開コマンドを受信する。MLDプロファイルがポリシーと関連付けられている場合、展開コマンドは検出サーバへのプロファイルの展開をもたらす。DLPポリシーがアクティブなDLPポリシーである場合、MLDプロファイルはアクティブとなり、ドキュメントを監視するのに即座に用いることができる。MLDプロファイルが以前展開されたバージョンを持っている場合、そのバージョンは、新たなバージョンが展開されるまでは、展開されたままとなることに留意されたい。MLDプロファイルの新たなバージョンを展開すると、より古いバージョンは置き換えられてしまう場合がある。

【００６２】

ブロック４４０では、MLマネージャは、ユーザがトレーニングデータセットに変更を加えるべきであると提案する。トレーニングデータセットのドキュメントをユーザがカテゴライズしていた場合、MLマネージャはトレーニングデータセットに追加されるべきドキュメントのカテゴリを特定することができる。例えば、特定のカテゴリのドキュメントが多数の偽陽性を起こした場合、MLマネージャは、そのカテゴリのドキュメントをトレーニングデータセットの陰性ドキュメントにもっと加えてみることを提案できる。

【００６３】

ブロック４４５では、MLマネージャが、トレーニングデータセット又はメモリ割り当て選択に関して変更が加えられたかを判断する。トレーニングデータセット又はメモリ割り当てに変更が加えられている場合、方法はブロック４２０へと戻る。それ以外の場合、方法は終了する。展開されたプロファイルは、展開プロファイルページにて表示されることができる。このページは、現在展開されているプロファイルについての表示をユーザに提供する。

【００６４】

図９は、本発明の１つの実施形態による、MLDプロファイル生成時／変更時におけるMLマネージャの様々な状態を示す状態図９００である。ユーザが、新たなMLDプロファイル生成せよとのコマンドを入力すると、MLマネージャは“新規”状態９０５に入り、これによりMLマネージャは一時的ワークスペース及び空のMLDプロファイルを生成させられる。“新規”状態９０５からは、MLマネージャは、“プロファイルを管理する”状態９１０に入ることができる。“プロファイルを管理する”状態９１０では、MLマネージャは、ユーザ入力に基づいてトレーニングデータセットにドキュメントを追加することができる。また、MLマネージャは、MLDプロファイルを以前の状態にロールバックして“新規”状態９０５に戻ることができ、又は“トレーニング”状態９１５へと進むことができる。以前生成されたMLDプロファイルが変更中の場合、MLマネージャは、“プロファイルを管理する”状態９１０から“アクセプテッド”状態９３０へと遷移することができる。

【００６５】

“トレーニング”状態９１５にある間は、MLマネージャはMLDプロファイルをトレーニングする。トレーニングがキャンセルされる又は他の理由により失敗した場合、MLマネージャは“トレーニングが失敗／キャンセルされた”状態９２０へと遷移する。ユーザによるアクノレジメントの後、MLマネージャは“プロファイルを管理する”状態９１０に復帰する。トレーニングが成功した場合、MLマネージャは“トレーニング済み”状態９２５に遷移する。その後、ユーザはMLDプロファイルを棄却してMLマネージャを“プロファイルを管理する”状態９１０に戻すか、又はMLDプロファイルをアクセプトしてMLマネージャを“アクセプテッド”状態９３０へと遷移させることができる。“アクセプテッド”状態からは、MLマネージャはMLDプロファイルを展開することができる。

【００６６】

図１０は、MLDプロファイルを生成すること及び既存のDLPポリシーへMLDプロファイルを展開することについての方法１０００についての１つの実施形態を示すフローチャートである。方法１０００は、ハードウェア（回路、専用のロジック等）、ソフトウェア（汎用コンピュータシステム又は専用機で実行されるもの）又は両者の組合せを含むことができる処理ロジックにより行われる。方法１０００は、図１のエンフォースメントサーバ１

２０上で実行されているMLマネージャ１３０のようなMLマネージャによって行われることができる。

【００６７】

図１０を参照するに、ブロック１００５では、MLマネージャがDLPポリシーによって生成されたインシデントレポート及び／又は履歴データを収集する。インシデントレポートは、非機密ドキュメントとして誤分類されたドキュメント及び／又は機密ドキュメントとして誤分類されたドキュメントを含むことができる。また、インシデントレポートは、機密ドキュメントとして正しく分類されたドキュメント及び／又は非機密ドキュメントとして正しく分類されたドキュメントを含むことができる。

【００６８】

ブロック１０１０では、MLマネージャは、インシデントレポート／履歴データからのドキュメントをMLDプロファイルのためのトレーニングデータセットに追加する。ブロック１０１５では、MLマネージャは、機械学習を用いてトレーニングデータセットを分析してMLDプロファイルをトレーニングする。これには、特徴セットを生成すること、分類モデルを生成すること、及び１以上のクオリティーメトリックをMLDプロファイルについて生成することを、含むことができる。ブロック１０２０では、MLマネージャは、DLPポリシーにMLDプロファイルを追加する。

【００６９】

方法１０００は、既存のDLPポリシーについてのインシデントを用いてMLDプロファイルをどのようにして生成するかを示す。したがって、MLマネージャは、方法１０００を行って、既存のDLPポリシーを改良して、従来は分類に失敗していたドキュメントについてドキュメントを機密又は非機密として分類できるようにすることができる。

【００７０】

図１１は、MLDプロファイルを含むDLPポリシーを用いてデータロスからコンピューティング装置を保護する方法１１００についての１つの実施形態を示すフローチャートである。方法１１００は、ハードウェア（回路、専用のロジック等）、ソフトウェア（汎用コンピュータシステム又は専用機で実行されるもの）又は両者の組合せを含むことができる処理ロジックにより行われる。方法１１００は、図１のエンドポイント装置１０２A上で実行されているDLPエージェント１０６のようなDLPエージェントにより行われることができる。方法１１００は、図１のエンドポイントサーバ１１５上で実行されているグローバルDLP検出エンジン１２２のようなグローバルDLP検出エンジンにより行われることができる。

【００７１】

図１１を参照するに、ブロック１１０５では、処理ロジックは、ドキュメントに対して操作を行うことについての要求を受信する。ブロック１１１０では、MLモジュールが、MLDプロファイルを用いてドキュメントを分析してドキュメントを分類する。ブロック１２２５では、処理ロジックが、ドキュメントファイルあ機密又は非機密として分類されたかを判断する。ドキュメントが機密として分類された場合、方法はブロック１３３０へと進み、DLPレスポンスルールにおいて指定された動作が行われ、また、インシデントレポートが生成される。これには、操作を阻止すること、インシデントレスポンスレポートを生成すること等が含まれ得る。ドキュメントが非機密と分類された場合、方法はブロック１１３５へと進み、操作が行われる。

【００７２】

図１２は、コンピュータシステム１２００として例示的に示された、本明細書中で説明した１以上の任意の手法をマシンに行わせるための命令のセットを実行することのできるマシンの概略図的表現を示す。他の実施形態では、マシンは、LAN、イントラネット、エクストラネット又はインターネットで、他のマシンに接続（例えば、ネットワーク）されることができる。マシンはクライアントサーバモデルにおけるサーバとして若しくはクライアントマシンとして又はピアツーピア（又は分散型）ネットワーク環境下でピアマシンとして作動できる。マシンは、パソコン（PC）、タブレットPC、セットトップ（STB）、

パーソナルデジタルアシスタント（PDA）、携帯電話、ウェブアプライアンス、サーバ、ネットワークルータ、スイッチ、ブリッジ又はマシンで行われるべき動作を指定する（シーケンシャル若しくはそうでない）命令のセットを実行できる他の任意のマシンであることができる。さらに、１つのマシンのみが図示されていても、“マシン”との用語は、本明細書中で説明した１以上の手法を行うための命令のセット（又は複数のセット）を個別に又は合同的に実行する任意のマシン群をも含むものとして解されるべきである。

【００７３】

例示的なコンピュータシステム１２００は、処理装置（プロセッサ）１２０２、メインメモリ１２０４（例えば、リードオンリメモリ（ROM）、フラッシュメモリ、シンクロナスDRAM（SDRAM）やRambus DRAM（RDRAM）等のダイナミックランダムアクセスメモリ（DRAM）等）、静的メモリ１２０６（例えば、フラッシュメモリ、スタティックランダムアクセスメモリ（SRAM）等）、及びデータ記憶装置１２１８を含み、これらは互いにバス１２０８を介して通信する。

【００７４】

プロセッサ１２０２は、１以上のマイクロプロセッサや中央処理装置等の汎用処理装置を表す。より具体的には、プロセッサ１２０２は複数命令セットコンピューティング（CISC）マイクロプロセッサ、縮小命令セットコンピューティング（RISC）マイクロプロセッサ、超長命令語（VLIW）マイクロプロセッサ若しくは他の命令セットを実装するプロセッサ又は命令セットを複数組み合わせて実装するプロセッサであることができる。プロセッサ１２０２は、１以上の特定用途向け集積回路（ASIC）、フィールドプログラマブルゲートアレイ（FPGA）、デジタルシグナルプロセッサ（DSP）、ネットワークプロセッサ等の特殊用途処理装置であることもできる。プロセッサ１２０２は、本明細書中で説明する操作及びステップを行うための命令１２２６を実行するように構成されている。

【００７５】

コンピュータシステム１２００は、更にネットワークインターフェース装置１２２２をさらに含むことができる。コンピュータシステム１２００はビデオディスプレイ装置１２１０（例えば、液晶ディスプレイ（LCD）又は陰極線管（CRT））、アルファニューメリック入力装置１２１２（例えば、キーボード）、カーソル制御装置１２１４（例えば、マウス）、及び信号生成装置１２２０（例えば、スピーカ）を含むこともできる。

【００７６】

データ記憶装置１２１８は、本明細書中で説明される手法及び機能の１以上の任意のものを実装する１以上の命令１２２６のセット（例えば、ソフトウェア）が格納されるコンピュータ可読記憶媒体１２２４を含むことができる。命令１２２６は、コンピュータシステム１２００による該命令の実行時において、完全に又は少なくとも部分的にメインメモリ１２０４上に及び／又はプロセッサ１２０２内に存在することもでき、メインメモリ１２０４及びプロセッサ１２０２はコンピュータ可読記憶媒体たり得る。命令１２２６は、ネットワークインターフェース装置１２２２を介してネットワーク１２７４上でさらに送受信されることもできる。

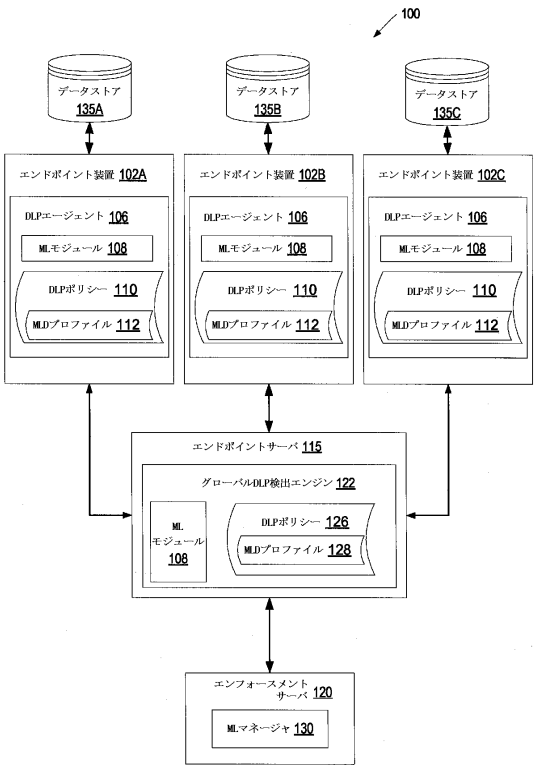
【００７７】

１つの実施形態では、命令１２２６は、図２のMLマネージャ２０５及び／又はMLマネージャをコールするメソッドを含むソフトウェアライブラリ等のMLマネージャのための命令を含む。例示的実施形態ではコンピュータ可読記憶媒体１２２４は単一媒体として示されているものの、“コンピュータ可読記憶媒体”との語は１以上の命令のセットを格納する単一の又は複数の媒体（例えば、集中型又は分散型のデータベース及び／又は関連するキャッシュ及びサーバ）を含むものと解されるべきである。“コンピュータ可読記憶媒体”との語は、また、機械による実行のための命令のセットを格納、エンコード又はキャリーすることのできる任意の媒体であり、本発明の手法の１以上の任意のものを機械に実行させるもの、を含むものと解されるべきである。したがって、“コンピュータ可読記憶媒体”との語はソリッドステートメモリ、光学媒体及び磁気媒体を含むもとの解されるべきであるがこれらには限定されない。

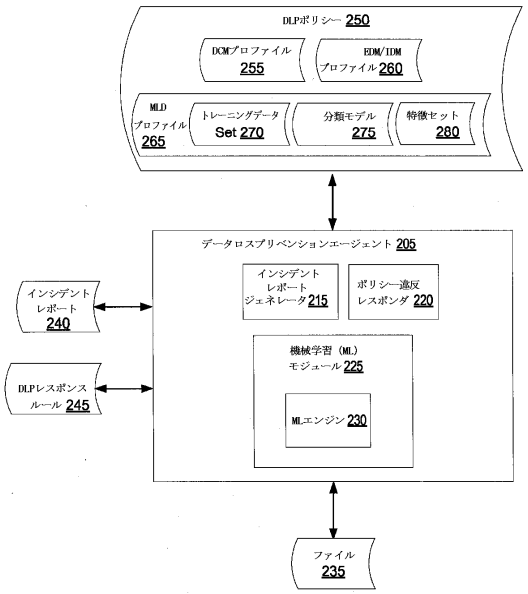
【００７８】

上記説明は例示的なものであり、限定的なものと解されないべきである。上述の説明を読んで理解した当業者には多くの他の実施形態が見えるであろう。したがって、本発明の範囲は、添付の請求項とそれらの請求項が享受すべきその等価物の全範囲を参照して決定されるべきである。

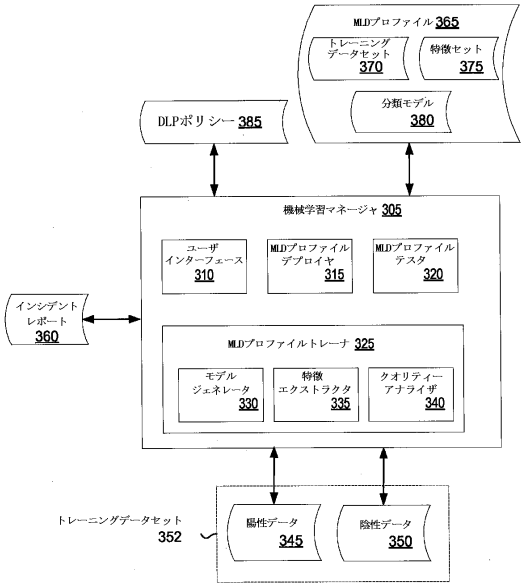
【図１】



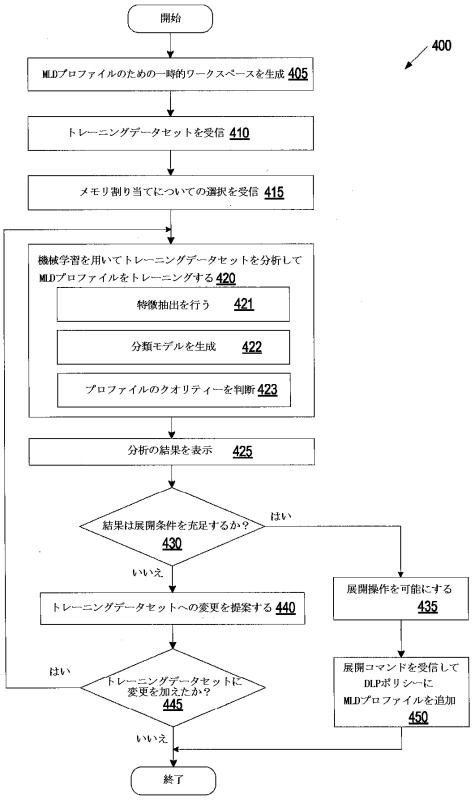
【図２】



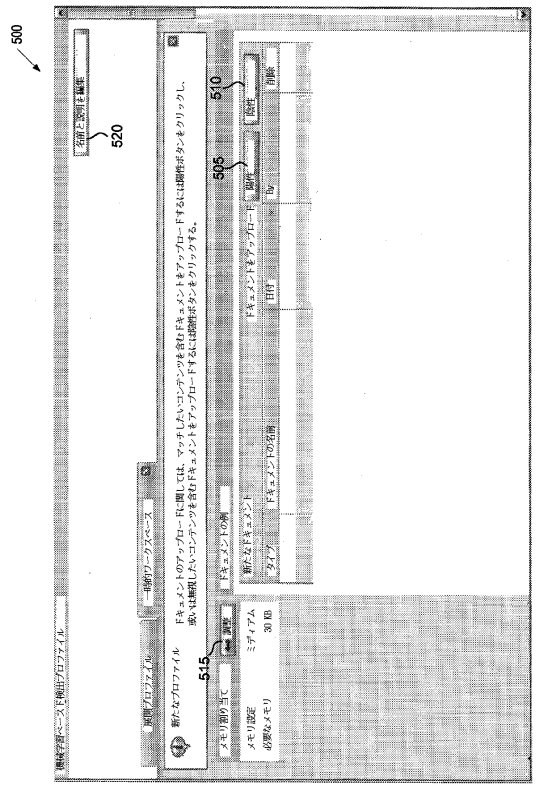
【図３】


MLDプロファイル 365
トレーニングデータセット 370
特徴セット 375
分類モデル 380
DLPポリシー 385
機械学習マネージャ 305
ユーザインターフェース 310
MLDプロファイルデプロイヤ 315
MLDプロファイルテスタ 320
インシデントレポート 360
MLDプロファイルトレーナ 325
モデルジェネレータ 330
特徴エクストラクタ 335
クオリティーアナライザ 340
トレーニングデータセット 352
属性データ 345
陰性データ 350

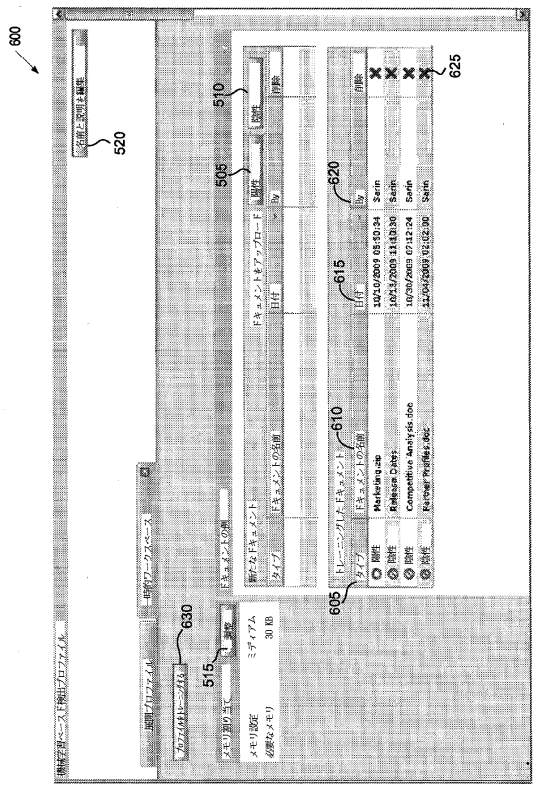
【図４】


400
開始
MLDプロファイルのための一時的ワークスペースを生成 405
トレーニングデータセットを受信 410
メモリ割り当てについての選択を受信 415
機械学習を用いてトレーニングデータセットを分析してMLDプロファイルをトレーニングする 420
特徴抽出を行う 421
分類モデルを生成 422
プロファイルのクオリティーを判断 423
分析の結果を表示 425
結果は展開条件を充足するか？ 430
はい
いいえ
トレーニングデータセットへの変更を提案する 440
展開操作を可能にする 435
トレーニングデータセットに変更を加えたか？ 445
はい
いいえ
展開コマンドを受信してDLPポリシーにMLDプロファイルを追加 450
終了

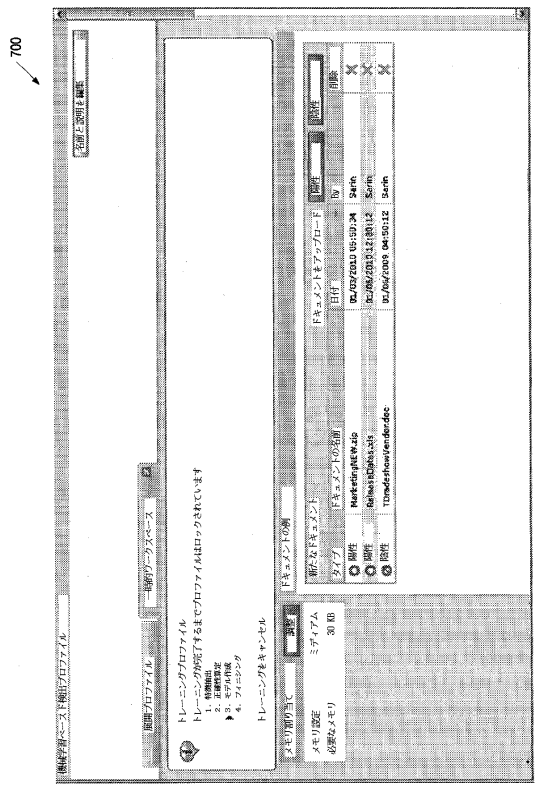
【図５】


500
520
510
505
515
新たなドキュメント
ドキュメントをアップロード
ドキュメントの例
メモリ設定
必要なメモリ
ミディアム
30 MB

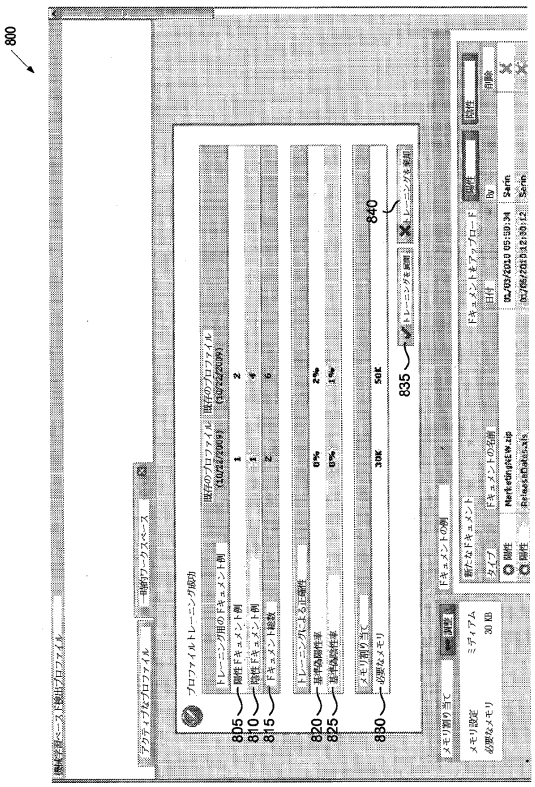
【図６】


600
520
510
505
515
630
605
610
615
620
625
新たなドキュメント
ドキュメントをアップロード
ドキュメントの例
メモリ設定
必要なメモリ
ミディアム
30 MB
トレーニングドキュメント
Marketing.xls
Release Dates
Competitive Analysis.xlsx
Partner Profiles.doc

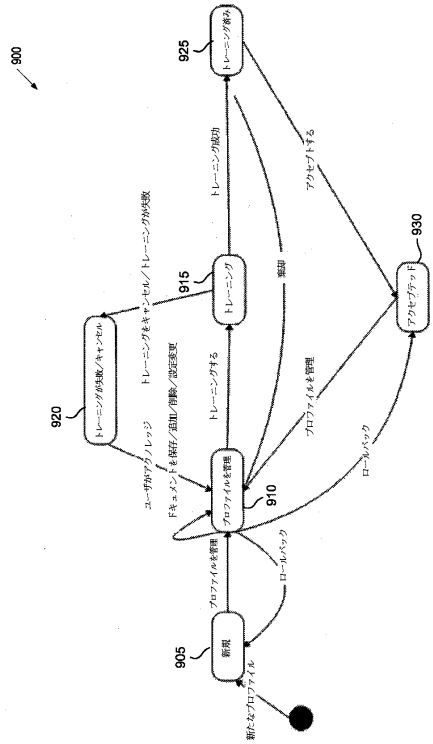
【図７】


700

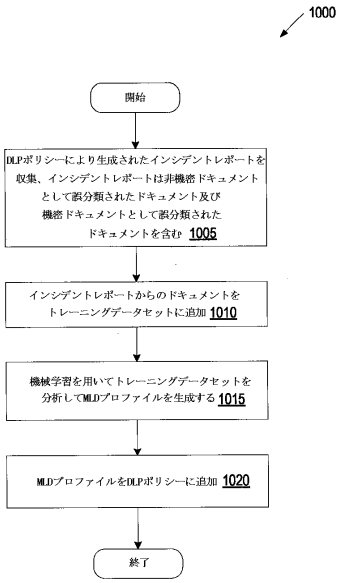
【図８】


800
805
810
815
820
825
830
835
840

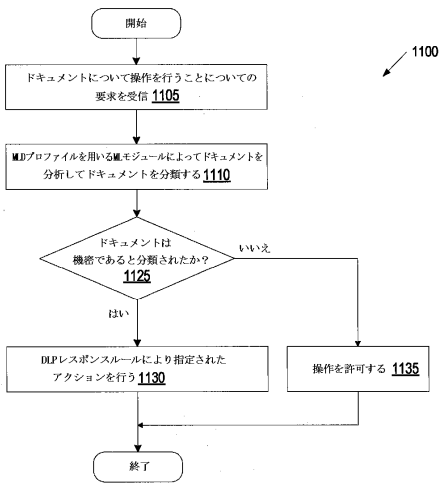
【図９】


900
905
910
915
920
925
930
新規
プロファイルを管理
トレーニング
トレーニング/キャンセル
トレーニング済み
アクセプテッド
新たなプロファイル
ロールバック
トレーニングを開始
トレーニングキャンセル/トレーニング失敗
トレーニング成功
アクセプトする
削除
プロファイルを管理
ロールバック
ユーザがアクノレッジ
ドキュメントを保存/追加/削除/設定変更
トレーニングする

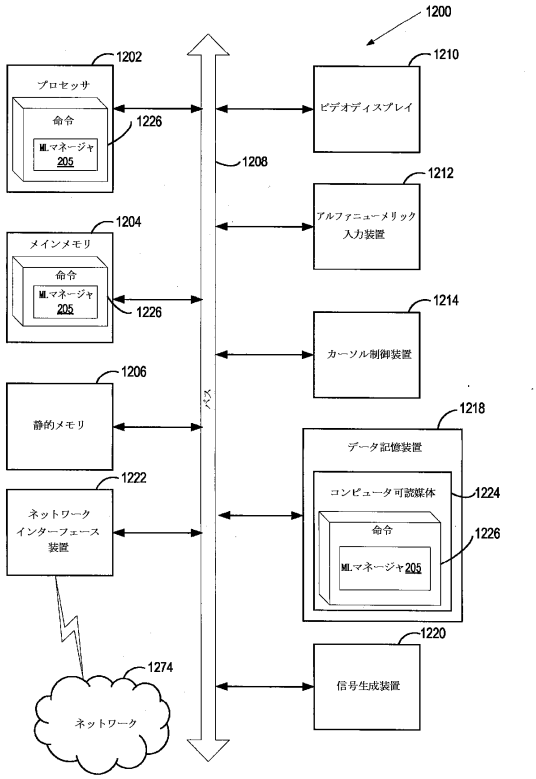
【図１０】


1000
開始
DLPポリシーにより生成されたインシデントレポートを収集、インシデントレポートは非機密ドキュメントとして誤分類されたドキュメント及び機密ドキュメントとして誤分類されたドキュメントを含む 1005
インシデントレポートからのドキュメントをトレーニングデータセットに追加 1010
機械学習を用いてトレーニングデータセットを分析してMLDプロファイルを生成する 1015
MLDプロファイルをDLPポリシーに追加 1020
終了

【図１１】


1100
開始
ドキュメントについて操作を行うことについての要求を受信 1105
MLプロファイルを用いるMLモジュールによってドキュメントを分析してドキュメントを分類する 1110
ドキュメントは機密であると分類されたか？ 1125
いいえ
はい
DLPレスポンスルールにより指定されたアクションを行う 1130
操作を許可する 1135
終了

【図１２】


1200
1202
プロセッサ
命令
MLマネージャ 205
1226
1208
バス
1204
メインメモリ
命令
MLマネージャ 205
1226
1206
静的メモリ
1222
ネットワークインターフェース装置
1274
ネットワーク
1210
ビデオディスプレイ
1212
アルファニューメリック入力装置
1214
カーソル制御装置
1218
データ記憶装置
コンピュータ可読媒体
1224
命令
MLマネージャ205
1226
1220
信号生成装置

フロントページの続き

(72)発明者 フィリップ ディコルポ
アメリカ合衆国 カリフォルニア州 ９４１１４ サンフランシスコ コーリングウッド ストリート ２９
(72)発明者 シタルクマル エス サワント
アメリカ合衆国 カリフォルニア州 ９４５３８ フレモント レッド ホーク サークル １４０１ アパートメント オー１０１
(72)発明者 サリー カウフマン
アメリカ合衆国 ワシントン ディーシー ノースウェスト ナンバー２ フィフティーンス ストリート １９１８
(72)発明者 アラン デール ガリンデズ
アメリカ合衆国 カリフォルニア州 ９４０２５ メンロー パーク メンロー オークス ドライブ ４８１－１／２
(72)発明者 スメッシュ ジャイスワル
インド国 プネー ４１１００６ カリヤニ ナガー サンシャイン コート フラット １８ ユニット ２
(72)発明者 アシシュ アガルワル
インド国 ウッタル プラデーシュ ２４７００１ サハーランプル ビハット ロード ブハグワッティ コロニー ハウス ナンバー １２０

審査官 多胡 滋

(56)参考文献 特開２０１０－１９８４９８（ＪＰ，Ａ）
特開平１１－３４４４５０（ＪＰ，Ａ）
山田剛良，迷惑メール対策の大本命 送信者認証のしくみ，日経ＮＥＴＷＯＲＫ，日本，日経ＢＰ社，２００４年 ８月２２日，第53号，pp.090-094
中村洋幸，外４名，ベイジアンフィルタに基づく研究者検索システムの開発，電子情報通信学会技術研究報告 ＰＲＭＵ２００７－１３５～１５７ パターン認識・メディア理解，日本，社団法人電子情報通信学会，２００７年１２月 ６日，第107巻，第384号，pp.7-12
奥山真一郎，Ｗｉｎｄｏｗｓサーバー環境の管理を効率化する期待のツールがこの春いよいよ登場，ＳＱＬ ＳＥＲＶＥＲ ｍａｇａｚｉｎｅ，日本，株式会社翔泳社，２００３年 ３月 １日，第7号，pp.40-44
柴田秀哉，外３名，機密メール検出における訓練用データ自動収集手法，[online]，電子情報通信学会データ工学研究専門委員会，２０１０年 ５月２５日，[2016年4月15日検索]，ＵＲＬ，http://db-event.jpn.org/deim2010/proceedings/files/B4-1.pdf
竹林知善，外３名，情報漏えい防止セキュリティ技術開発への取組み，ＦＵＪＩＴＳＵ，日本，富士通株式会社，２００９年 ９月１０日，第60巻，第5号，pp.444-450
中台慎二，外１名，サポートベクターマシンを用いた事例ベース障害検出，電子情報通信学会技術研究報告 ＩＣＭ２００８－４０～４９ 情報通信マネジメント，日本，社団法人電子情報通信学会，２００８年１１月 ６日，第108巻，第288号，pp.1-6
学習型メールフィルタでスパムメールを撃退 ＰＯＰＦｉｌｅ，Ｌｉｎｕｘ ｍａｇａｚｉｎｅ，日本，株式会社アスキー，２００４年 １月 １日，第6巻，第1号，pp.120-121

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｎ ９９／００
Ｇ０６Ｆ １７／３０
Ｇ０６Ｆ ２１／６２